AD ____________________

MIPR NO: 95MM5560

TITLE: Vulnerability of Female Produced Speech in Operational Environments

PRINCIPAL INVESTIGATOR(S): Charles W. Nixon, Linda J. Morris, Angela R. McCavitt, Mark P. McDaniel, Timothy R. Anderson, Richard L. McKinley, Donald G. Yeager

CONTRACTING ORGANIZATION: Armstrong Laboratory
Wright-Patterson AFB, Ohio 45433-7901

REPORT DATE: August 31, 1995

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE: August 31, 1995

3. REPORT TYPE AND DATES COVERED: Annual (5 Dec 94 to 1 Aug 95)

4. TITLE AND SUBTITLE: Vulnerability of Female Produced Speech in Operational Environments

5. FUNDING NUMBERS: 95MM5560

6. AUTHOR(S): Charles W. Nixon, Linda J. Morris, Angela R. McCavitt, Mark P. McDaniel, Timothy R. Anderson, Richard L. McKinley, Donald G. Yeager

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES): Armstrong Laboratory, Crew Systems Directorate
Wright-Patterson AFB, Ohio 45433-7901

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES): US Army Medical Research and Materiel Command
Fort Detrick MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

19960524 026

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT: Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The role of women in combat environments has expanded dramatically. The accommodation of females in these previously male dominated environments is vital to fully successful mission accomplishment. Effective voice communications are critical to successful strategic and tactical operations. Current aircraft communications systems and components were optimized for male voice characteristics. Based on laboratory experience and knowledge of the basic characteristics of male and female acoustic speech, many questions exist concerning possible differential effects of operational variables on female speech. This study investigates the impact of the following factors on the perception of female speech: (a) the different spectra of operational noise environments, (b) the response characteristics of noise-cancelling microphones, (c) digital encoding and decoding, and (d) accuracy of automatic speech recognition systems.

14. SUBJECT TERMS: Female Speech, Speech Perception, Female vs Male Speech, Speech in Noise, Speech Intelligibility, Voice Communications
Defense Women's Health Research Program

15. NUMBER OF PAGES: 48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

CWN _X_ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Charles W Nixon 8/28/95

PI - Signature Date

TABLE OF CONTENTS

PROGRESS REPORT 1 SEPTEMBER 1995

Vulnerability of Female Produced Speech in Operational Environments

"No other essential activity in aircraft operations is as vulnerable to failure through human error and performance limitations as spoken communications." Monan (1986) cited in reference 20.

INTRODUCTION

A research program has been initiated to examine the perception of female speech produced in operational environments by listeners in operational environments. Emphasis is on female aviators and selected systems and conditions that are elements of typical aircraft voice communication systems. Speech performance is being measured in the cockpit noise environments of four different types of aircraft, with noise-cancelling microphones, with digital speech coders and decoders, and with automatic speech recognition systems (voice controllers). Female speech performance is being evaluated relative to male speech perception and to performance criteria that indicate the relative effectiveness of the female speech under operational conditions.

Vigilance is essential to ensure the effective voice communications that are critical to successful strategic and tactical operations. Numerous system, operator, and environmental factors can degrade effective communications to marginal or unacceptable levels. The basic designs and the performance of current aircraft audio communication systems have remained the same for several decades and need to be upgraded to incorporate current technologies. Special speech vocoders and encryptors dismantle and later reconstruct the acoustic speech signal that is often less robust and more vulnerable to noise than the original signal. Noise can directly degrade speech communications by interfering with or masking the speech signal and it can indirectly degrade it further by causing temporary and permanent noise induced hearing loss in the aviators. It can also interfere with the operation of voice recognition or voice control systems which are unable to extract the aviator speech signal commands from the noise. These factors have been dealt with for a long time without full success. They must receive continual attention to maintain effective voice communications and avert difficult and life threatening operational situations caused by the inability to communicate.

A situation is emerging that introduces a new factor that may, or may not, decrease the effectiveness of voice communications. Women are already flying high performance aircraft and their increasing presence in the cockpits and crew stations of Department of Defense (DoD) strategic and tactical aircraft is assured. Current aircraft audio communication systems and components were optimized for male voice characteristics and may not fully accommodate the female voice. Current knowledge of the perception of female speech, particularly in the harsh environments of military aviation, is not sufficient to allow reliable estimates of female speech performance in the cockpit environment. This Air Force study will seek the information necessary to identify significant differences, if present, in the perception of female and male speech. Differences that would prevent female speech from communicating effectively in current weapon systems must be addressed. Difficulties with the perception of female speech would affect all aviators.

Fundamentals of Human Speech

In the study of the human voice, the variability in characteristics from talker to talker is a dominant feature. Consequently, when the acoustic speech records of groups of talkers are analyzed, different acoustic spectra can be obtained. However, a basic feature of the speech sounds and the frequency regions in which their maximum amplitudes occur is that they are about the same and are generally independent of the talker. It is this basic feature that allows the acoustic characteristics of speech to be studied systematically.

The perception of female and male speech is essentially equivalent under almost all typical living conditions (ranging from a whisper in church to a shout at the playground); however, recognizable differences are obvious. The bases for these differences are associated with the acoustic speech signals generated by the male and by the female talker. The acoustic components of the female speech signal are almost always higher in frequency than those of the male. The fundamental frequency of the average female voice is about 250 Hz and of the average male voice is about 125 Hz. The speech spectra for average male and female speech are similar, with the female spectrum higher than the male spectrum by about 5 to 10 dB at 4000 Hz and above and lower by about 12 dB at 125 Hz and below. The high frequencies of the vowel sounds are 5 to 15 percent higher, the mid-high frequencies 5 to 25 percent higher, and the low frequencies up to 35 percent higher than the corresponding frequencies in the male voice. The average speech power for males is 34 microwatts and for females is 18 microwatts which corresponds to a difference of about 3 dB at conversational speech level (8).

In addition to gender differences, the acoustic features of an individual's speech are continuously changing for various voluntary and involuntary reasons. A talker may emphasize segments of speech, alter speech rate and level, shout, talk during physical exertion, and speak with emotion. Speaking in a raised voice, in order to be understood in the presence of a background noise or to talk to a distant listener, requires increased vocal effort. The accompanying muscle strain usually causes an increase in the pitch of the voice, and can cause vocal cord fatigue over time. These changes also influence differences between female and male speech.

Human speech is very robust and is easily understood in many distorted forms. Accents, incorrect pronunciation, foreign dialect, speech compression, peak clipping, and digital coding and decoding of the speech signals may sound unnatural, yet be very intelligible. In spite of the robust nature of speech and its ability to be universally understood, it is subject to degradation under various conditions. Degradation can be caused by unfavorable speech-to-noise ratios, distortions, communications channels, terminal equipments that include microphones and earphones, workload, stress, and the individual talker and/or listener. Factors which degrade speech communications in military applications must be identified and their impact on operations evaluated.

Operational Variables and Communications Effectiveness

Crew stations in military aircraft contain many factors with the potential to decrease voice communications effectiveness even though the stations have been designed to optimize performance. Perhaps the most pervasive factor at these stations is acoustic noise. Noise is caused by numerous sources including vehicle propulsion systems, environmental systems, life support systems, weapons fire, and air turbulence, as well as the voice communication system itself. One of the primary effects of the noise is masking of the voice communication signals. In general, when the level of the noise in the frequency region of the speech sounds exceeds the level of the speech, communications are degraded. The ratio of the level of the speech to the level of the noise (signal-to-noise ratio, SNR) provides an estimate of the level of the speech performance; the higher the ratio, the better the speech performance. Also, some learning is involved in becoming an effective communicator in noise environments; understanding speech in noise improves with practice (16). Persons experienced with communicating over military systems in noise usually perform very well.

Noise

Over the past two decades, voice communications research has been conducted with the human-in-the-loop in the Bioacoustics and Biocommunications Branch of the Air Force Armstrong Laboratory. The major research facilities within the branch contain ten communication stations; consequently the standard procedure used in investigations is to simultaneously utilize the ten experimental subjects. The panels of trained subjects, over the two decades of research, have consisted of five males and five females. Although research during that period did not focus on female speech, some studies involved comparisons of female-male speech performance. In general, these measurements and observations have revealed that the performance of female speech has, in most instances, been lower or less effective than male speech under the same conditions. The current study is concerned with the systematic measurement and evaluation of some of these differences.

In a 1991 summary of a study by Backs and Walrath, it was stated that "...under conditions of high noise stress, female speakers were less intelligible than males..." (3). In an earlier study of voice communications in simulated cockpit noise, a systematic difference was measured between the intelligibility of male and female talkers. In levels of noise at 95 dB and below, there was essentially no difference in intelligibility. At levels of noise at 105 dB the female talkers were seven percent less intelligible than males and at 115 dB the difference increased to ten percent. The differences at both the 105 dB and 115 dB levels of noise were significant at the 95 percent level of confidence (15). A study of positive pressure breathing effects on speech intelligibility was conducted in aircraft noise at levels of 65 dB, 95 dB, 105 dB and 115 dB. The results of this study reflect these same general observations of effects of noise on female and male speech intelligibility. The differences between the perception of female and male speech increase as the levels of the noise increase, with the female speech becoming less intelligible. The maximum decreases of female speech intelligibility occur at the highest levels of sound. In that study, only the differences at the 115 dB level of noise were significant at the 95 percent level of confidence (19).

Noise-Cancelling Microphones

The adverse effect of noise on speech transmission promoted the development of noise-cancelling microphones, again based on the male voice. Aviators now use two general types of noise-cancelling microphones, a "kiss-to-talk" microphone (lips touch microphone for maximum performance) for applications such as oxygen masks, and a boom type microphone for headsets and helmets worn by personnel in environs such as tanks and transport aircraft. The speech intelligibility of female aviators using either type of microphone has not been measured. An early evaluation of female and male speech intelligibility was conducted with the M-101 microphone, a former Air Force standard microphone. The M-101 microphone was compared to a modified M-101, that had been reduced 50 percent in thickness to improve its fit in an oxygen mask. The intelligibility of the female and male speech measured in a 95 dB level of noise was essentially the same for the standard M-101 microphone. However, the speech intelligibility of the male voice increased eight percent with the "thin M-101" microphone whereas the female speech increase was only three percent. These differences were not statistically significant (14). However, the lower female speech intelligibility usually observed under noise conditions was present. Also, the five percent difference measured in the 105 dB level of noise would be expected to be somewhat larger in higher levels of noise.

Speech Coders

Speech coders have been added to military voice communication systems to increase and maintain the reliable transfer of information. Speech coders convert the analog speech signal to digital units which are transmitted to the receiving station where they are converted back to speech. During this process some of the analog speech signal is lost; the amount of the signal that is lost is a major factor that determines the quality of the vocoded speech. The effectiveness of the vocoding process and the amount of information lost depends on the characteristics of the analog-to-digital-to-analog conversion system.

Earlier research with three versions of the Department of Defense standard Linear Predictive Coding (LPC-10) speech coder demonstrated that its intelligibility was poor and that it was vulnerable to voice communication degradation due to acoustic noise at the listener (17). These data were revisited as part of the current study and performance of the male and female speech was extracted and examined. Female speech in high performance aircraft and in combat was not an issue at the time of the original study. Although the sample size was very small (two female and three male talkers), the average intelligibility with the three LPC-10 vocoders at four levels of noise was essentially the same for males and females. On the basis of other research efforts, involving four levels of noise, it had been predicted that the female speech would be less intelligible rather than equal to the male speech. Consequently, the study and instrumentation were re-examined and it was discovered that the gain of the speech signal available to the subjects (who individually adjust the gain for their own headset systems) had been limited. This undiscovered limitation prohibited the subject from increasing the gain of her/his individual intercommunication system to improve speech communications. Without the limitation on gain, it is assumed that the male speech perception would have been better than the female speech perception; however, whether or not the difference would be significant cannot be estimated from the available information. The intelligibility of female speech processed by the standard LPC-10 vocoder and perceived in noise environments must be determined empirically.

Automatic Speech Recognition

Automatic speech recognition or voice control systems are very effective when properly trained to recognize the talker and when used in relatively quiet environments. However, the success of these systems has generally been limited in high level noise environments because of their inability to discriminate the components of the acoustic noise signal from the acoustic speech signal. Even though the speech recognition system has been taught to recognize a talker ("memorizes" speech components during training), it can be fooled to interpret components of the noise as elements of the speech, resulting in incorrect recognition. The aircraft cockpit is a particularly hostile environment for voice control systems, yet it is one that can derive substantial benefit from the successful implementation of voice control.

Current noise cancellation microphones reduce the level of the noise as a function of frequency, but they do not eliminate the noise. Also, the acoustics inside the oxygen mask are further complicated by sounds, such as the aviator breathing, as well as valve noise during each respiration cycle, added to the external noise that has reached the inside of the oxygen mask (18). In spite of these persistent problems, state-of-the-art speaker-dependent voice control systems (also one speaker-independent system) have been designed specifically for the cockpit environment (22). Some of the manufacturers of these systems report word recognition accuracy of over 80 percent for connected digits and over 95 percent for words spoken as two-word phrases in 90 dB of noise (90 dB is well below many operational noise environments). These systems generally function with limited vocabularies and with substantial talker training. Speaker-independent systems do not require training. Recognition accuracy also varies with the talker.

Voice control technology is already present, to a limited degree, in several aircraft. Utilization of voice control in the noisy cockpit is expected to increase; however, no major breakthroughs in voice control technology appear to be on the horizon. There is no database of the recognition of female speech by voice control systems in cockpit-like noise environments. Knowledge of factors such as the lower acoustic power of the female voice and its reduced intelligibility in higher noise levels indicates that voice control with the female voice in operational noise environments must be evaluated.

Voice Warning

An indirectly related area of female speech perception is that of voice advisories and voice warning signals. The initial installation of a voice warning system in an Air Force military aircraft was in 1961 when an audio tape system was installed in the fleet of B-58 Hustler aircraft. Early evaluations indicated that aviators felt that voice warnings contributed to flight safety, that pilot reaction time was improved, that warning recognition time improved by six to nine seconds, and that the female voice was the preferred warning signal (9). Voice warning systems have been evaluated in terms of aviator preferences (includes voice quality) and of quantitative metrics such as accuracy of response, reaction time, and speech intelligibility. Although aviators are relatively firm in their initial judgments for particular voice characteristics, their preferences tend to change with their continued exposure to those voices in operational situations (23).

Contrary to early beliefs, subsequent research has demonstrated that the female voice is not the preferred warning signal and that it usually ranks low in terms of both preference and quantitative metrics. It is reported that the male voice had greater accuracy and a shorter response time than the female voice in a 105 dB noise environment (9). Another report measured no differences in the intelligibility of male and female voice warnings in a noise environment of 95 dB (23). Also, a distinctive, mechanical quality voice that could be recognized in a background of human voices was preferred over either female or male voice warnings. New technologies provide a great deal of flexibility, and at relatively low cost, for the generation of voice warning systems. It is not unreasonable to expect that successful systems might use a variety of voices, both human and synthesized, to create a menu driven system that allows the aviator to select the suite of voice and auditory warning signals that she/he believes will provide the best performance. It is unclear if there is any relationship between the low ranking of female speech as a voice warning signal and its "lower than male speech" intelligibility under various conditions such as high levels of noise.

RESEARCH OBJECTIVES

Dramatic transitions are underway with the acceptance of females as military aviators in a profession formerly occupied only by males. The total aviation environment and all related facilities and equipments were designed and evaluated for the male. Numerous efforts are underway attempting to identify those situations in the aviation environment with which the female is not fully compatible and to evaluate their impact on performance and safety. Voice communication and its effectiveness under a variety of different operational situations and circumstances is one of the most important areas under investigation. It is almost universally accepted that in-flight voice communications must be free of errors. In a report on civil aviation Billings and Cheany (1981) state in reference 20, "Problems in the transfer of information between the aviation system were noted in over 70 percent of 28,000 reports submitted by pilots and air traffic controllers...during a 5-year period 1976-1981. These problems are related primarily to voice communications..." It would be unusual to find a reader who does not know of some situation in which a breakdown of communication has resulted in an unacceptable consequence.

Under normal conditions, the understanding of female and male speech is equivalent even though there are obvious differences in the acoustic speech signals. In situations in which factors such as noise degrade speech, female speech intelligibility is reduced more than that of males. These differences in speech associated with gender, and the reduction in intelligibility, tend to increase with increasing levels of noise. When this reduction in intelligibility reaches certain levels, speech communication is no longer effective.

The research objectives of this study are to quantify the differences between the perception of female and male speech relative to those factors in operational situations that influence voice communications, to determine whether the reductions in speech performance are or are not significant relative to operational environments, and to propose actions to minimize significant effects, where feasible.

The specific questions selected for investigation are, to what extent is the perception of female (and male) speech affected by:

(a) the different cockpit noise environments (spectra) of four operational aircraft,

(b) the response characteristics of standard military noise-cancelling microphones,

(c) digital encoding and decoding of the speech signals with the DoD standard LPC-10 vocoder, and

(d) the recognition accuracy of voice controllers/automatic speech recognition systems for female speech.

APPROACH

The acoustic speech signal differences based on gender have not been systematically investigated in environments emulating operational conditions that include standard communication systems and equipments in realistic acoustic environments. This study initiates such an effort; however, the large number of these environments and the time required to emulate all of the operational conditions of interest are prohibitive for a one-year study. The research team considered a variety of questions relative to their possible impact on the mission, time frame of the study, and laboratory resources that could immediately be brought to bear on the issues. It was agreed that the four proposed phases of the study would evaluate communication performance in a reasonable representation of operational conditions and of speech communication technologies.

The initial phase of the study examined speech performance in typical aircraft cockpit noises (5, 6, 7, 10). Four different aircraft noise spectra were selected that represent the range of cockpit noise environments in which female aviators are found. These cockpit noise environments include the low frequency spectra of the C-130E aircraft and MH-53 helicopter, the relatively flat spectrum (up to 4000 Hz) of the C-141B, and the higher frequency spectrum of the F-15A tactical aircraft. The noise spectra shown in Figure 1 represent the levels of noise experienced in the fixed-wing aircraft cockpit positions during normal cruise flight conditions and during hover at 50 feet in the helicopter aircraft. The flight deck, as well as other crew locations, can experience levels of noise much higher than those observed during cruise and hover. In Phase I, speech performance was measured for each of the aircraft in four different levels of the cockpit noise spectra. In Phase II, the relative effectiveness of the current standard noise-cancelling microphones was examined in the same noise environments employed in Phase I.

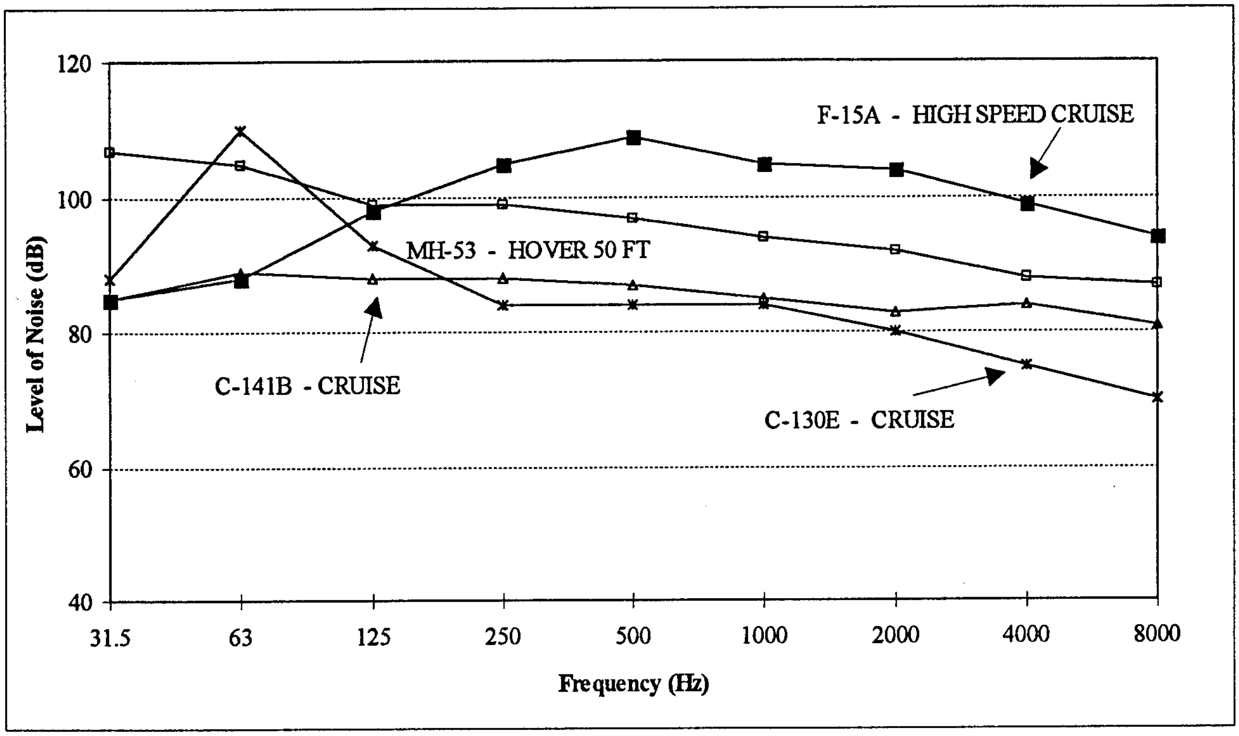


Figure 1: Aircraft cockpit noise spectra.

The intelligibility of the male and female speech processed by the Department of Defense standard LPC-10 speech coder and a high quality speech coder (Continuously Variable Slope Delta modulation system, CVSD) was examined in Phase III. As noted earlier, the coder converts the analog speech signal to a digital signal that is transmitted to the receiver where it is reconverted to speech. Some of the speech signal is lost in this conversion process. Phase III examined the robustness of the reconstructed female speech in the presence of the four aircraft noise conditions of Phase I.

Control of critical operations in the cockpit by voice commands requires highly accurate recognition systems. In Phase IV, the recognition accuracy of female and male speech by two different automatic speech recognition (ASR) systems is evaluated in cockpit noise environments. Voice control is already present in cockpits and it is expected to extend to more aircraft and require greater numbers of commands per aircraft. Some of the better ASR systems are reported to obtain 90 to 95 percent accuracy in noise levels of about 90 dB. However, accuracy can fall off sharply as the level of the cockpit noise increases. Recognition accuracy by ASR systems of male and female speech in aircraft noise has not been reported. Female speech voice control will be empirically examined in this study.

Criterion Measure

The criterion measure for Phases I, II, and III is the percent correct intelligibility of the Modified Rhyme Test (MRT)(10). The MRT is the test of choice for evaluating the performance of military communication systems and equipments. The materials consist of word lists that are equivalent in intelligibility. Each list contains 50 monosyllable words in the form of consonant-vowel-consonant. During the investigation, the talker speaks each of the 50 test words in a list in the carrier phrase, "Number ___, you will mark ___ please." The listeners select the word they believe was spoken by the talker from a set of six words that rhyme with the spoken word. The listener's intelligibility score is the percent correct adjusted for correct answers obtained by guessing ($2.4 \times$ number correct - 20). The score for the experimental condition is the average of the scores of the ten listeners. The MRT does not require extensive training of subjects and is relatively simple to administer, score, and evaluate. The measurement of speech intelligibility in this study was accomplished in accordance with the American National Standard, S3.2-1989, Method for Measuring the Intelligibility of Speech Over Communication Systems (2).

The criterion measure for Phase IV is recognition accuracy of voice commands created to initiate actions in the cockpit environment. The test procedure does not require human subjects to respond to the speech materials. The talkers will speak the commands into the automatic speech recognition systems which are used as the listeners.

Performance Criteria

The Bioacoustics and Biocommunications Branch, at Wright-Patterson AFB OH, maintains a vigorous research program in all aspects of voice communications effectiveness. The laboratory uses dedicated facilities designed to evaluate all the system, operator, and environmental variables that can degrade voice communications. The data and experiences obtained using the Modified Rhyme Test, the standardized procedures, and the Voice Communication Research and Evaluation System (VOCRES) laboratory facilities (Figure 2) revealed a high relationship with performance in the operational situation. For example, a head-mounted bone conduction microphone, designed for Air/Sea Rescue applications, exhibited performance that failed the laboratory performance criteria. Development continued, but the microphone subsequently failed the Operational Test and Evaluation program. A different microphone used in a new low-profile oxygen mask also failed the speech communications performance criteria, but was still provided to operational fighter pilots. Field performance was so poor that the aviators were prohibited from flying with that microphone. Conversely, active noise reduction headsets, crew helmets, and new noise-cancelling microphones are examples of equipments that were acceptable under the performance criteria and remain highly successful in the operational situation. These examples verify the relationship between the Biocommunications Laboratory performance criteria and actual performance under operational conditions. Consequently, a set of speech intelligibility performance criteria measured in the laboratory was adopted several years ago and it continues to be utilized to successfully estimate and predict corresponding performance in the field.

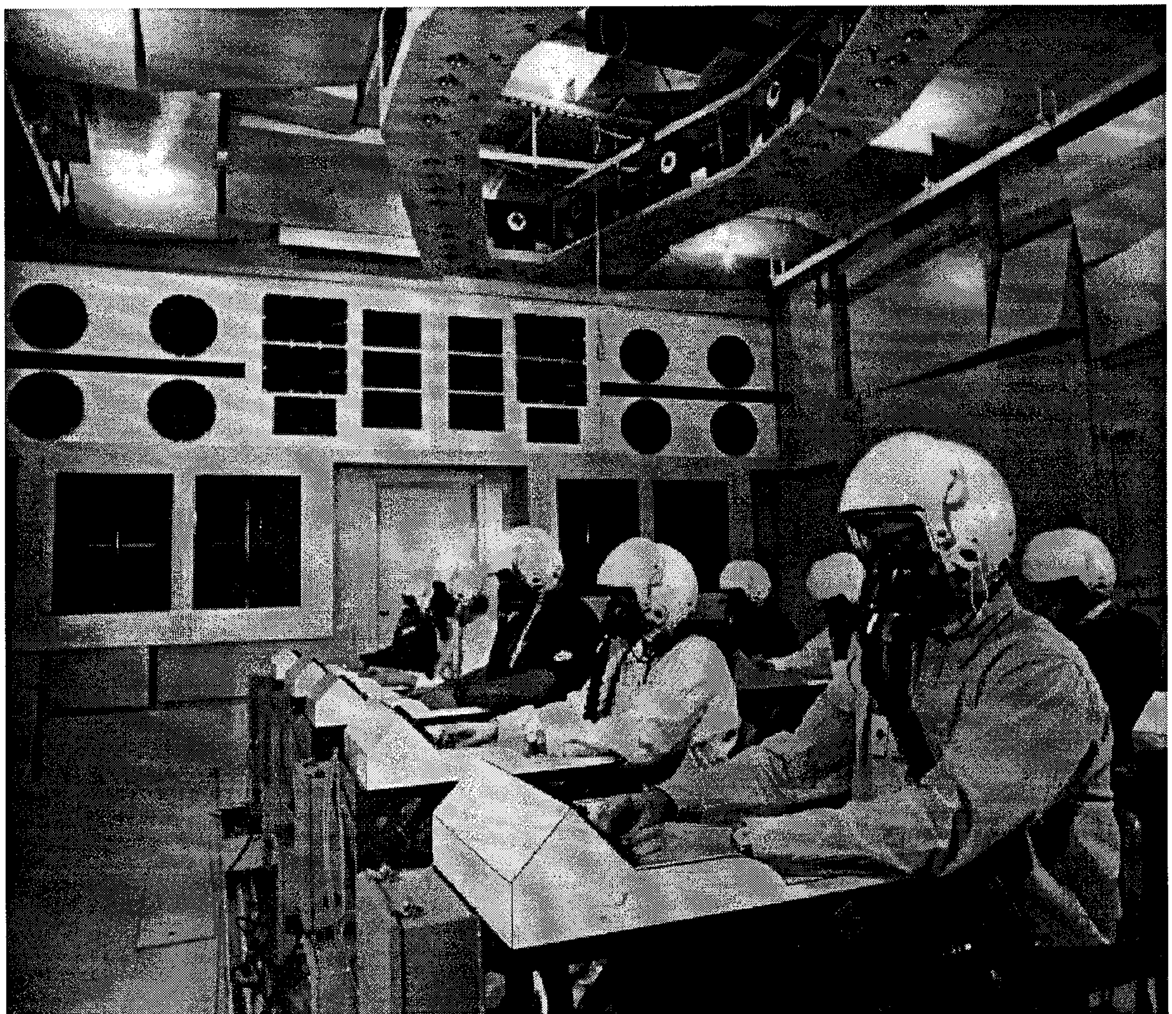


Figure 2: Voice Communications Research and Evaluation System (VOCRES).

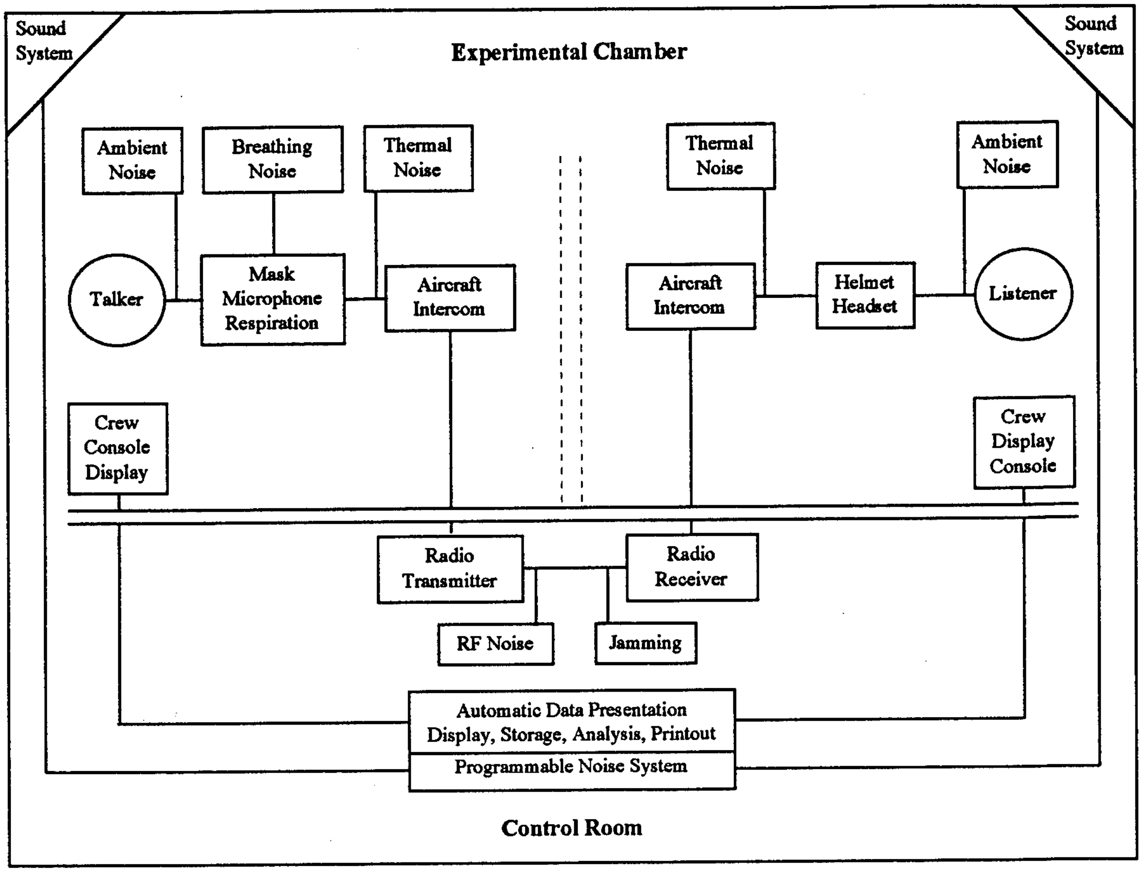


Figure 3: Configuration of the VOCRES facility.

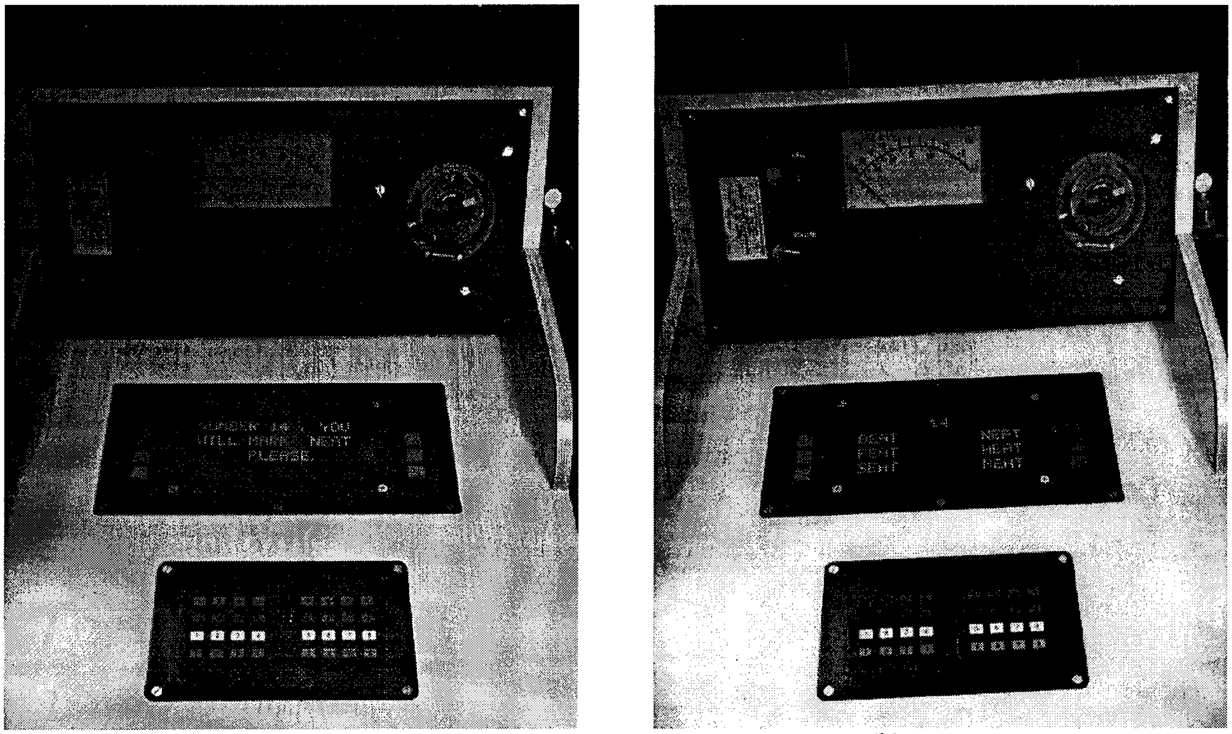
The performance criteria predict that systems, components, and materials displaying speech intelligibility performance of about 70 percent correct (MRT) and below are typically unacceptable in corresponding operational applications. Those with performance in the range from about 70 percent to 80 percent are considered marginal and their success in the field depends on the specific conditions under which they are utilized. Marginal situations would include those for which there is ample time to repeat messages to achieve understanding. Those exhibiting intelligibility performance of about 80 percent correct and above are fully acceptable under operational conditions. Speech performance measured under the various conditions in this study of female speech perception was examined in terms of these performance guidelines. These guidelines have been very useful in many situations, such as those in which differences in the measured speech intelligibility are statistically significant but the amount of the difference is so small that it is not meaningful in field situations.

Subjects

This investigation utilized human subjects who were experienced in voice communications research in the laboratory. All were recruited from the general population and were paid an hourly rate for their participation. All subjects spoke midwestern American English and none exhibited a noticeable accent, dialect, or speech problem. Twenty adult subjects, ten males and ten females, participated throughout all phases of the study. All subjects participated as talkers and a subset of ten subjects (five male and five female) comprised the listening panel. Subjects exhibited normal hearing sensitivity and middle ear function, as verified by pure tone audiometry and tympanometry, prior to participation in the study. Monitoring audiometry was performed biweekly throughout the study to insure no individuals incurred a hearing threshold shift. A communication headset/helmet was custom fit to each subject and worn by that subject throughout the study. Sound attenuation of each unit (Appendix B) was measured while worn by each subject to insure that she/he received adequate hearing protection during the study (1).

Facilities and Equipment

The study was conducted in the VOCRES facility in the Armstrong Laboratory Crew Systems Directorate (13). This voice communication research system located in a large reverberation chamber contains the operator, system, and environmental variables known to most directly affect voice communication effectiveness (Figure 2). VOCRES consists of a central processing unit that controls the experimental sessions and the subject stations (Figure 3). The facility contains ten individual automated communication stations which provide simultaneous measurement of all test subjects. Each station is equipped with an alphanumeric light emitting diode (LED) display, a subject response unit consisting of special keyboards for entering performance responses to the central processing unit, and a large volume unit (VU) meter that indicates voice level of the speech produced by the talker at that station (Figure 4). Each station contains an Air Force standard helmet/headset, air respiration system with oxygen mask, and aircraft intercommunication system. Aircraft radios, electronic warfare instrumentation, secure speech units, speech vocoders, and a wide-passband research intercommunication system are also imbedded in the VOCRES. In Phases I and II, an additional communication station was located inside VOCRES to accommodate the individual talker in the same noise environment as the ten-member listening panel.



(a) (b)

Figure 4: (a) VOCRES talker station. (b) VOCRES listener station.

VOCRES also contains a programmable sound system that can generate high intensity levels of noise in the laboratory. These high level noise environments can be created with laboratory equipment or can utilize noise data which were previously recorded in crew locations in aircraft to provide high quality emulations of operational environs. The overall system allows the accurate recreation in the laboratory of essentially any operational voice communication situation and noise environment. Air Force standard headsets, helmets, and microphones used for this study are those currently found in operational aircraft. The headset/helmet systems used are listed with the appropriate aircraft in Table 1.

Aircraft	Headset/Helmet System	Microphone
C-130	H-157 Headset	M-87
C-141	H-157 Headset	M-87
F-15	HGU-55P Helmet with MBU/P oxygen mask	M-169
MH-53	SPH-4AF Helmet	M-87

Table 1: Phase I - Aircraft, headset/helmet, and microphone combinations tested

Two digital speech coding systems, called vocoders, were selected to process the speech signals in Phase III. These systems use the natural speech signal that is segmented, processed, coded, and later decoded to provide the speech output. The vocoders utilized in this study are the DoD standard LPC-10 and the CVSD systems. LPC-10 predicts the current speech sample from a linear combination of previous speech samples. It is based on the voicing, pitch, reflections, and amplitude of the speech. This information is processed into standard LPC format. LPC is reported to be vulnerable to noise (17,18). CVSD uses an algorithm that codes only the difference between one speech sample and the next sample. Basically, the difference is coded and used to predict the next speech sample in this ongoing process. CVSD is robust in noise.

The two automatic speech recognition systems to be employed in Phase IV are the ITT VRS-1290 and the IBM VoiceType. These systems represent two different technologies for continuous speech recognition. The ITT VRS-1290 is a speaker-dependent speech recognition system. This means that each individual talker must train the recognition system to recognize her/his speech production. This is accomplished by the talker speaking each vocabulary word into the system a number of times until the system indicates that the criterion level of recognition has been attained. The ITT system has a vocabulary of 500 words and uses the Dynamic Time Warping (DTW) technology to perform its pattern recognition and matching at the word level. This system uses special purpose hardware on a personal computer (PC). An earlier version of this system was flown in an F-15 aircraft during the mid 1980s (24). The current version of this system has been tested and flown in an Army helicopter (11).

The IBM VoiceType is a speaker-independent system. This means that it requires no specific training of the system to recognize the individual talker. The IBM system has a vocabulary of over 30,000 words and uses the Hidden Markov Model (HMM) technology to represent words with sub-word units called phonemes. This process enables additional words to be added to the system recognition vocabulary by adding the sequence of phonemes for the new word to the dictionary. This system runs on a personal computer without special purpose hardware, except for an analog-to-digital converter. This system has been evaluated in laboratory environments (23).

Experimental Systems Calibration and Measurement

Prior to data collection, all equipment was calibrated to ensure reliability, conformity to specifications, and accuracy. Earphone outputs were measured for the H-157 headset, and for the HGU-55P and SPH-4AF helmet communication units. Each earcup was placed on an artificial ear with a flat plate coupler and 2 volts rms were applied at frequencies of 125, 250, 500, 1k, 2k, 4k, and 8k Hz. Output values were logged and compared; differences between the outputs of the two earphones in a headset unit did not exceed 5 dB. Frequency responses were obtained from measurements of the voltage output of each M-87, M-162, and M-169 noise-cancelling microphone by placing the microphone 1/4" away from an artificial voice with an output level of 95 dB using a Brüel and Kjær 4134 reference microphone (Appendix A). One microphone of each type, representative of the measured average response of that type of microphone, was selected for use in the experiment.

The VOCRES facility was calibrated by passing eight pure tones at octave spacings from 100 Hz to 6300 Hz through the system for analyses by an audio analyzer. The speech calibration frequency was 1000 Hz. Distortion and acoustic noise at the headsets of each station were within specifications, background noise was minimized, and VU meters were adjusted to provide appropriate visual feedback of voice volume to the talker at each station. Each of the ten stations was characterized by collecting frequency response data for the headphone and microphone.

GENERAL PROCEDURES

All data were collected with both the talker and listeners in the same noise environments. As previously noted, the experimental design required the measurement of the perception of the speech of twenty talkers by a panel of ten listeners. Twenty talkers were selected to expand the applicability of the data and findings of the study. Experience with voice communications in noise environments has revealed greater variance among the speech of groups of talkers than among listeners (18).

The C-130E, C-141B, F-15A, and MH-53 operational aircraft noise spectra were chosen for this study because they are representative of aircraft which are currently open to female aircrews and potentially vulnerable environments for female speech. The four noise conditions studied are representative of the typical range of noise spectra found at the pilot-copilot positions of the selected aircraft. Specifically, the four operational noise levels chosen for each aircraft consisted of an ambient noise condition of 66 dB and aircraft noise presented at 95 dB, 105 dB, and 115 dB.

During data collection, each member of the ten subject listening panel was seated at an experimental test station and one of the twenty talkers was seated at the remote experimental test station in the VOCRES facility. Each subject, each of the listeners, and the talker were equipped with the custom fit headset or helmet corresponding to the experimental condition being evaluated (see Table 1). For each experimental run, the word list appeared on the LED display in front of the talker, one word at a time. The talker read each word, after which each member of the listening panel selected the word she/he believed was spoken from the list of six rhyming words on the LED display by pressing the response button adjacent to that word. Data from each of the ten stations were sent simultaneously to a computer which calculated each listener's score for a specific talker for each experimental run. Data collection for each phase of the study followed this procedure for each experimental run in all noise spectra and levels investigated. The study was conducted in a series of four phases in which specific variables were investigated at each phase.

EXPERIMENTAL PHASES

Phase I

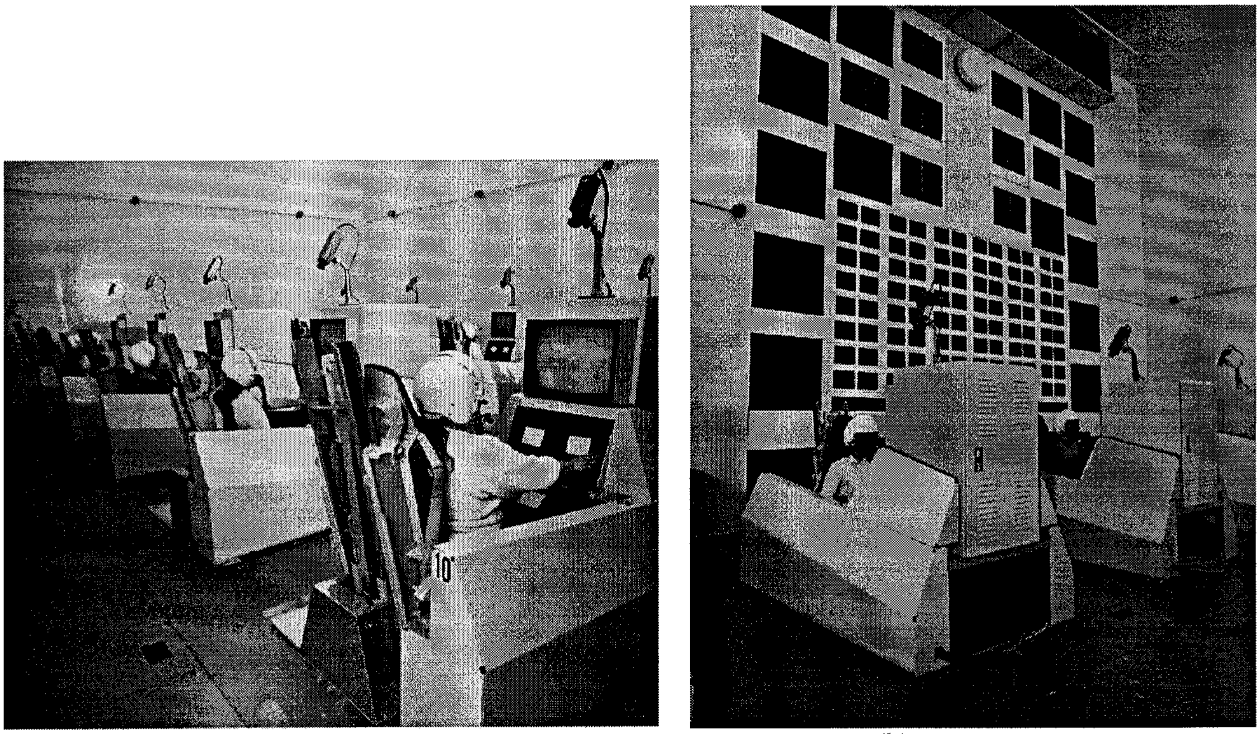
Phase I examined the influence of the spectrum and the level of four aircraft cockpit noises on the intelligibility of female and male speech. The three independent variables of subject, spectrum of noise, and level of noise were randomized to minimize effects such as variations in the repeat trials, subject differences, and learning. The dependent variable was percent correct speech intelligibility on the MRT. The operational noise spectra and levels previously noted were selected to identify potential areas requiring enhancements of female produced speech perceived by others in various noise environments. Phase I data were collected under a total of 320 conditions: four spectra × four levels for each spectrum × twenty subjects.

Phase II

The independent variables investigated in Phase II were noise-cancelling microphones, noise spectrum, and noise level; the dependent variable was speech intelligibility. Two standard noise-cancelling microphones used for this phase were the M-87 boom microphone and the M-162 microphone. Intelligibility of male and female produced speech was measured using the M-162 microphone in three noise spectra: C-130, C-141, and MH-53, each at four noise levels. Data collected on the M-87 microphone in Phase I were extracted for analysis in Phase II. The F-15 spectrum was not used in this phase because it requires the use of a helmet with an oxygen mask. The M-169 noise-cancelling microphone contained in the oxygen mask is the only microphone appropriate to use with the oxygen mask. Therefore, experimentation in Phase II included 240 total conditions: three spectra × four levels for each spectrum × twenty talkers for one microphone (M-87 microphone data were collected in Phase I).

Phase III

Phase III is being conducted to evaluate the effect of digital coding of speech signals on speech intelligibility. Speech signals are encoded and decoded using the DoD standard LPC-10 and the CVSD vocoder. Speech intelligibility performance is being evaluated with each system in the four noise spectra and four noise levels previously discussed. For this phase, the configuration of the experimental stations has been varied slightly to better emulate the operational environment in which digital coding devices are used. The remote talker station is placed in the Performance and Communication Research and Technology (PACRAT) facility, a facility capable of generating noise spectra and levels identical to those used in VOCRES. PACRAT contains all of the features of VOCRES, plus, task loading features. The ten stations in PACRAT are emulations of fighter aircraft cockpits and employ simultaneous dynamic performance tasks to load and overload the speech signals to determine their robustness (Figure 5). The talker in each experimental run is seated in the PACRAT facility in the same noise spectrum and level as the ten listeners seated the VOCRES facility. The speech signal is transmitted from the remote talker over phone lines via modem, coded and decoded by the system being used, and then received by the listeners who respond in the same manner as in all previous experimental phases. Phase III includes a total of 640 experimental conditions: four spectra × four levels × two coder-vocoders × twenty talkers.



(a) (b)

Figure 5: (a) PACRAT individual stations in the foreground. (b) PACRAT sound system in background.

Phase IV

The purpose of Phase IV of this study is to measure the recognition accuracy of female and male speech using two state-of-the-art automatic speech recognition (ASR) systems in two noise spectra (C-130E and F-15A) in each of the four levels of noise used in previous phases. The two fundamentally different systems that will be used are the ITT VRS-1290 and the IBM VoiceType speech recognition systems. The vocabulary chosen for use with these systems is a vocabulary currently being used in a joint Air Force-NASA in-flight study of voice control in the OV-10 aircraft. The headset/helmet and microphone combinations worn by the talkers for the C-130E and the F-15A conditions in the earlier phases of the study will also be worn by the talkers in Phase IV (see Table 1). As in previous phases, twenty subjects will be used as talkers; however, instead of the ten subject listening panel the two ASR systems will be used as the listeners. A total of 320 conditions will be investigated in this phase: two spectra × four levels of each spectrum × two ASR systems × twenty talkers.

RESULTS

Measurement data are provided for Phases I and II in this progress report. Data are comprised of measurements of speech intelligibility of ten male and ten female talkers as perceived by a panel of ten listeners (five male and five female). The responses of the individual subjects were averaged for each experimental condition. Means and standard deviations were calculated and differences among the means were evaluated using standard statistical paired t-tests at the 95 percent confidence level.

In the following data the criterion measure, average percent correct intelligibility, of the female speech is below that of the male speech in all conditions. Although the differences are relatively small and they range from about one percent to ten percent, at no time do the values for the female talkers equal or exceed those of the male talkers.

Data were treated by measures of central tendency and variance with emphasis on the average differences between the means of the samples. The statistical significance of the differences between the means of the matched pairs (female and male) was determined by calculating the t-score and comparing it with the criterion t-value corresponding to the 95 percent confidence level (4). The calculated t-scores for each pair indicate the number of standard deviations separating the two means. If the t-score is greater than the criterion t-value at the 95 percent confidence level, the difference between the paired means is considered to be statistically significant. However, the difference may not be considered operationally significant.

Phase I

Aircraft Cockpit Noise Spectra

The average intelligibility scores are summarized for the female and male subjects for each aircraft at the four levels of noise. The data are shown in graphical form in Figures 6 through 9 and in tabular form in Tables 2 through 5. The vertical bars on the figures represent plus and minus one standard deviation. Those differences between means that are statistically significant at the 95 percent level of confidence are circled on the graphs and are boxed in the tables.

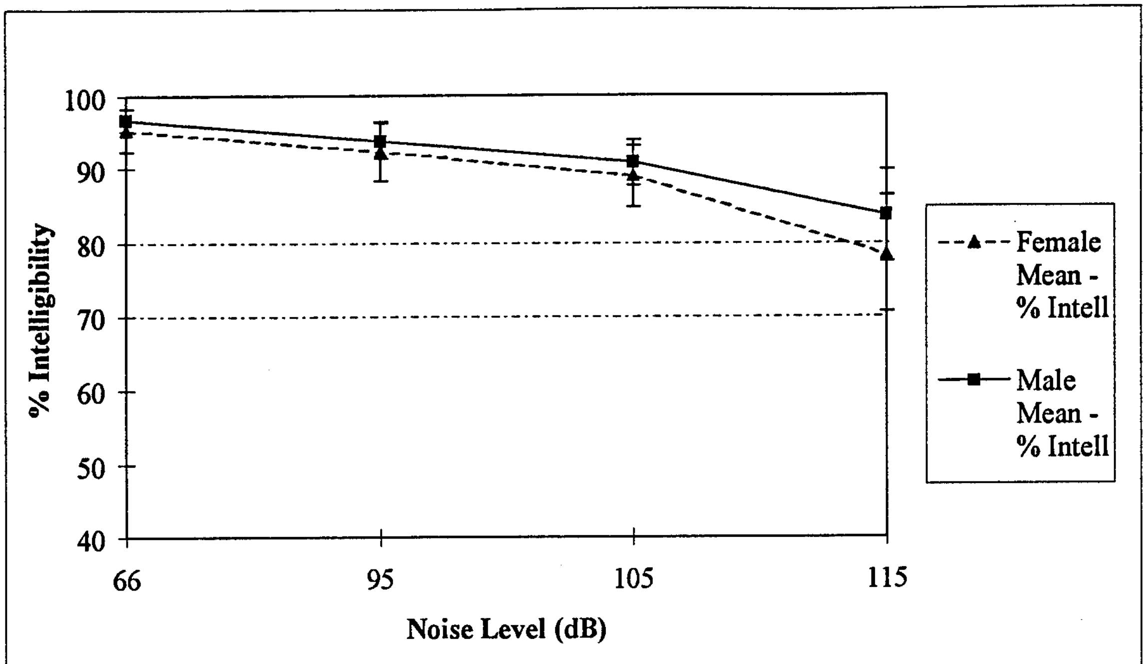


Figure 6: Phase I - Male versus female intelligibility using C-130 spectrum, H-157 headset, and the M-87 microphone.

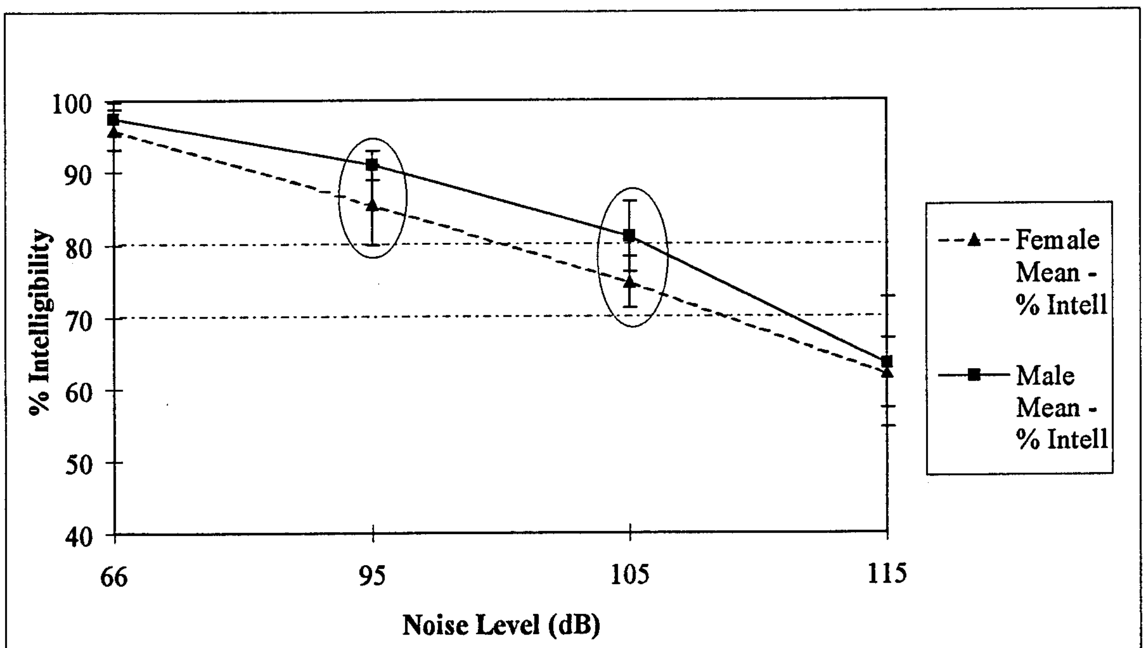


Figure 7: Phase I - Male versus female intelligibility using C-141 spectrum, H-157 headset, and the M-87 microphone.

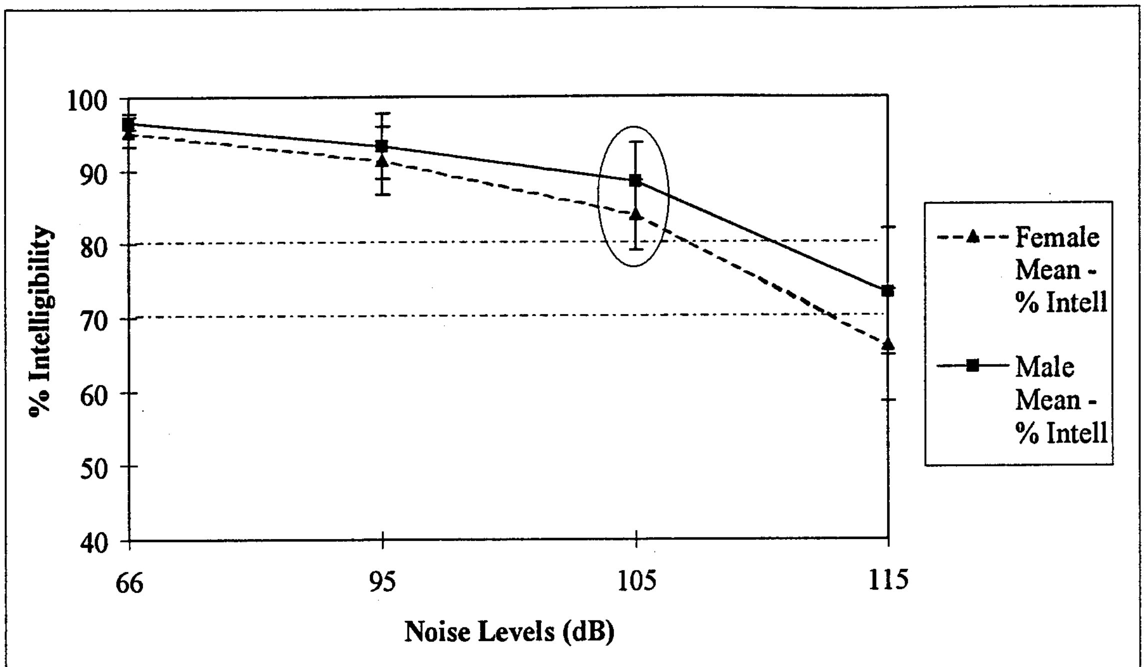


Figure 8: Phase I - Male versus female intelligibility using F-15 spectrum, HGU-55P helmet with MBU/P oxygen mask, and the M-169 microphone.

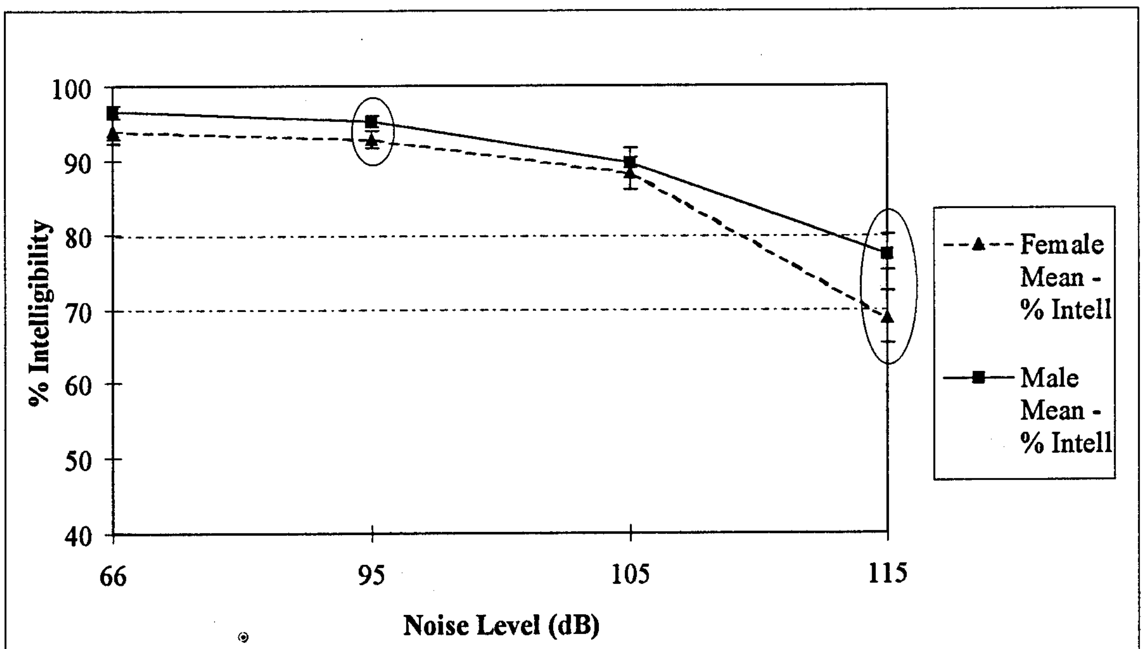


Figure 9: Phase I - Male versus female intelligibility using MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.

	66 dB	95 dB	105 dB	115 dB
Female - % avg. intelligibility ± standard deviation	95.4 ± 2.4	92.4 ± 4.0	89.0 ± 4.1	78.5 ± 7.9
Male - % avg. intelligibility ± standard deviation	96.7 ± 1.6	93.8 ± 2.5	90.8 ± 3.1	83.6 ± 6.2
Difference in Means	-1.3	-1.4	-1.8	-5.1
T-score	-1.29	-0.95	-1.1	-1.6

Table 2: Phase I - Male versus female intelligibility with C-130 spectrum, H-157 headset, and the M-87 microphone.

	66 dB	95 dB	105 dB	115 dB
Female - % avg. intelligibility ± standard deviation	96.0 ± 2.8	85.5 ± 5.6	74.9 ± 3.5	62.2 ± 4.8
Male - % avg. intelligibility ± standard deviation	97.5 ± 2.27	91.0 ± 2.1	81.1 ± 4.9	63.7 ± 9.0
Difference in Means	-1.5	-5.5	-6.2	-1.5
T-score	-1.29	-2.85	-3.26	-0.45

Table 3: Phase I - Male versus female intelligibility with C-141 spectrum, H-157 headset, and the M-87 microphone.

	66 dB	95 dB	105 dB	115 dB
Female - % avg. intelligibility ± standard deviation	95.3 ± 2.0	91.4 ± 4.7	83.9 ± 4.7	66.1 ± 7.5
Male - % avg. intelligibility ± standard deviation	96.6 ± 1.1	93.2 ± 4.4	88.5 ± 5.1	73.4 ± 8.5
Difference in Means	-1.3	-1.8	-4.6	-7.3
T-score	-1.89	-0.92	-2.10	-2.02

Table 4: Phase I - Male versus female intelligibility with F-15 spectrum, HGU-55P helmet with MBU/P oxygen mask, and the M-169 microphone.

	66 dB	95 dB	105 dB	115 dB
Female - % avg. intelligibility ± standard deviation	93.9 ± 4.1	92.8 ± 2.6	88.2 ± 5.2	68.9 ± 8.3
Male - % avg. intelligibility ± standard deviation	96.5 ± 2.1	95.3 ± 1.6	89.7 ± 4.8	77.3 ± 6.5
Difference in Means	-2.6	-2.5	-1.5	-8.4
T-score	-1.76	-2.62	-0.66	-2.53

Table 5: Phase I - Male versus female intelligibility with MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.

The aircraft noise spectra were examined to quantify the female speech performance in the noise environments of these aircraft in which female aviators are found. A matched experimental design was not implemented because three different headset-helmet-communication systems were worn by the subjects in the four aircraft noise spectra. The effectiveness of the personal equipment systems interacts with the noise spectra and levels to influence the speech intelligibility. These interactions were not examined in this study.

The frequency range of standard Air Force voice communication systems is approximately from 300 Hz to 3500 Hz. Noise spectra with substantial energy in this speech frequency region, or slightly below, are most effective in masking the speech signal. In the flat spectrum of the ambient noise at 66 dB, the speech signal was not masked and the intelligibility was essentially the same for all ambient conditions. The individual aircraft spectra were not presented in the ambient conditions; however, the subjects did wear the personal equipment items utilized in the respective aircraft measurements. Average intelligibility of male speech was 97 to 98 percent correct and for female speech was 94 to 96 percent. Even under these ideal conditions, 100 percent average intelligibility was not achieved.

The data in Figure 1 represent in-flight cruise conditions for which the spectra and level differ substantially among aircraft. In this study, the experimental conditions presented all the spectra at the same four fixed overall sound pressure levels (OASPL). This was done to include the range of levels found in almost all operational aircraft, to allow comparisons among aircraft types, as well as to measure reductions in speech performance as levels of noise spectra were increased for the individual aircraft.

The influence of spectra on speech performance can be compared for the C-130E and C-141B conditions because experimental subjects wore the same headset-microphone communications equipment in both sets of measurements. The only difference between the experimental conditions was noise spectrum. The comparison is also of interest because the speech performance of both males and females was best in the C-130E and poorest in the C-141B. Speech performance was acceptable in all measured conditions for the C-130E and was unacceptable for both male and female speech at the highest level C-141B spectrum.

The noise spectrum of the C-141B is very flat with a slight rolloff starting at about 4000 Hz. The C-130E spectrum has a high peak around 63 Hz that is more than 15 dB greater than the next highest octave band level in the spectrum. The C-130E spectrum rolls off at about 5 dB per octave starting around 1000 Hz in the central region of the passband of the voice communication equipment. The C-130E overall level is determined by the peak level of 111 dB; the levels of the other octave bands are so far below the peak that they make no contribution to the overall level. When the two spectra are at the same overall sound pressure level, the C-141 spectrum is higher than the C-130E spectrum in all bands except 63 Hz where it is less. The C-130E is the less effective masker of the two because of the lower levels in almost all bands and the rolloff starting at 1000 Hz.

The decreases in intelligibility due to increases in level vary with aircraft spectrum. Also, the amounts of the decreases become larger at the higher levels of noise. For the C-130E, the decrease in intelligibility is three percent less at 105 dB than at 95 dB, and seven to 10 percent less at 115 than at 105 dB. The C-141B intelligibility is 10 to 11 percent less at 105 dB than at 95 dB and 13 to 17 percent higher at 115 dB than at 105 dB. These decreases in intelligibility are approximately the same for male and female speech except at the 115 dB, C-141B condition where the decrease for females is larger and the 115 dB, MH-53 where it is smaller.

C-130E Aircraft

Perception of the male and female speech is essentially the same at the 105 dB level of noise and below with only a 5 percent difference at 115 dB (Figure 6). None of the differences is statistically significant. Both male and female speech are around the 90 percent correct region and above at noise levels of 105 dB and below. At 115 dB, the accuracy ranges from 79 percent correct for females and 84 percent correct for the males; both are acceptable. The overall level of the noise of the C-130 during maximum endurance cruise is about 111 dB in the flight crew compartment and a maximum level of 115 dB at one of the other crew stations (5). Speech perception in the crew compartment during cruise (111 dB) should be about 90 percent correct with the lowest intelligibility at any measured location in the aircraft of about 80 percent correct for the female. Voice communication conditions in this aircraft, for female and male talkers, are considered acceptable.

C-141B Aircraft

The speech intelligibility of both males and females was vulnerable to this noise spectrum, dropping in mean intelligibility almost 40 percent from the ambient to the 115 dB noise condition (Figure 7). The mean differences between genders at both the 95 dB and 105 dB noise conditions were statistically significant at the 95 percent level of confidence. Both female and male speech were acceptable at the 95 dB level, at 105 dB male speech is acceptable and female speech is marginal, and both were not acceptable at the 115 dB level. Assuming that the relatively linear function shown by the graph is reliable, the extrapolated percent correct intelligibility at 100 dB should be almost 80 percent for the female and acceptable; it should be higher at lower levels of noise. The overall level of the noise between the pilot and copilot on the C-141A is 96 dB with a worst case condition of 117 dB during taxi with four engines at taxi power and 111 dB during climb to 3000 feet (6).

F-15A Aircraft

Speech perception decreases and the differences between female and male mean speech intelligibility increase in the F-15A as the level of the noise increases (Figure 8). The only statistically significant difference between the mean values occurred at the 105 dB noise condition which was acceptable for both speech conditions. At the 115 dB level of noise, the male speech was marginal and the female speech unacceptable. The overall sound pressure level of the F-15A cockpit noise during cruise was about 110 dB and during a high speed run it was about 115 dB (7). The data suggest that female speech perception is marginal to unacceptable in the high noise environments of these two flight conditions and that the male operates in the marginal region at 115 dB. It is presumed that experienced aviators compensate to maintain communications for marginal situations when the maximum levels of noise are encountered. However, improvement is required for female speech to be understood by other aviators in the 110 dB - 115 dB levels of noise.

MH-53 Helicopter

The mean intelligibility response curves are similar for the MH-53 helicopter (Figure 9) and the F-15 fighter aircraft (Figure 8) with the scores in the helicopter noise slightly better (10). Statistically significant differences between male and female speech perception occurred at the 95 dB and 115 dB noise conditions. The small difference of about 2.5 percent at the 95 dB noise condition is statistically significant because the standard deviations are very small. The mean difference at the 115 dB level of noise is about 8 percent. The noise spectra of the MH-53 and the C-130 vehicles are very similar except for the peak at 63 Hz in the C-130 spectrum. The maximum level of the noise between the pilot and copilot during cruise is 111 dB, while under maximum cruise it is 115 dB. The speech perception of both female and male is acceptable at all except the 115 dB condition. At 115 dB, male speech is in the marginal region, close to the acceptable range. The female speech is a little below the marginal region and must continue to be considered unacceptable. Improvement in female speech perception is required in these high level noise environments for good recognition by other aviators.

Phase II

Noise-Cancelling Microphones

The basic conditions in Phase I in which the M-87 noise-cancelling microphone was used were repeated in Phase II with the M-162 noise-cancelling microphone. These two sets of data (Phase I M-87 microphone and Phase II M-162 microphone) were compared to evaluate the relative effectiveness in noise of the microphones with female and male produced speech. The two independent variables of aircraft noise spectrum and level of the noise were randomized to minimize effects due to uncontrolled variance. The dependent variable was percent correct speech intelligibility on the MRT. The M-169 oxygen mask noise-cancelling microphone was not included in this evaluation; since there is no alternative mask microphone available, the M-169 data collected in Phase I represent its performance in the spectra and levels of the noises of interest.

The average speech intelligibility for the M-162 microphone in the various levels of the aircraft spectra are shown in graphical form in Figures 10 through 12 and in tabular form in Tables 6 through 8. No statistically significant differences between female and male speech, which were essentially identical, were observed with the M-162 microphone. All performance was acceptable, according to the performance criteria, except for the 115 dB noise condition for the C-141 aircraft which was unacceptable for both male and female talkers.

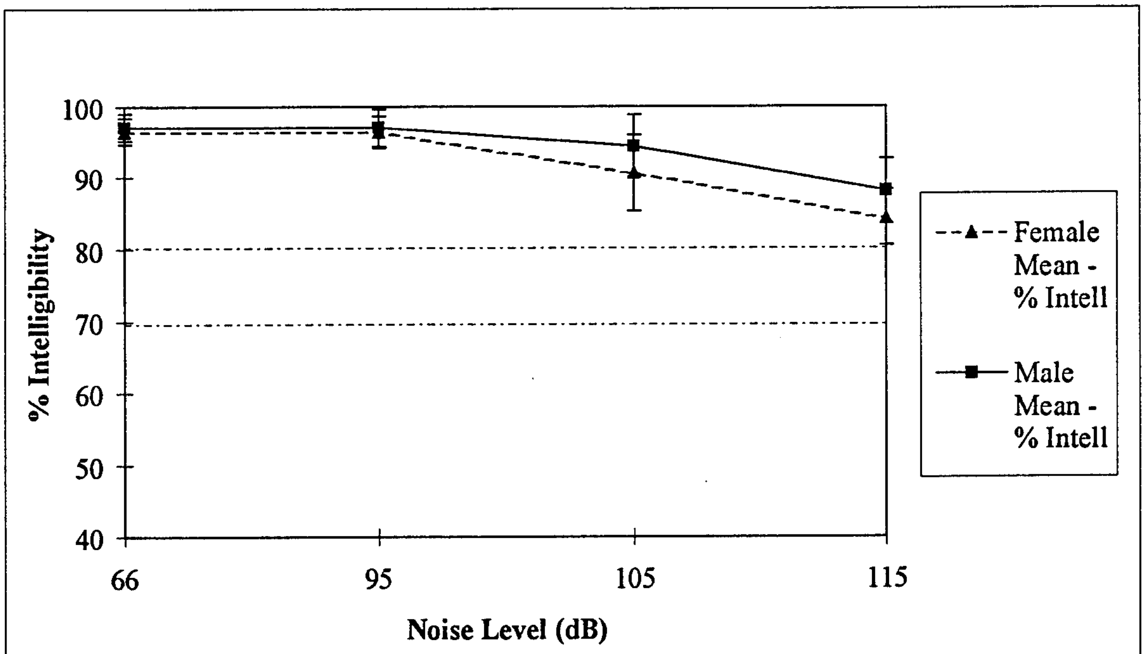


Figure 10: Phase II - Male versus female intelligibility with C-130 spectrum, H-157 headset, and the M-162 microphone.

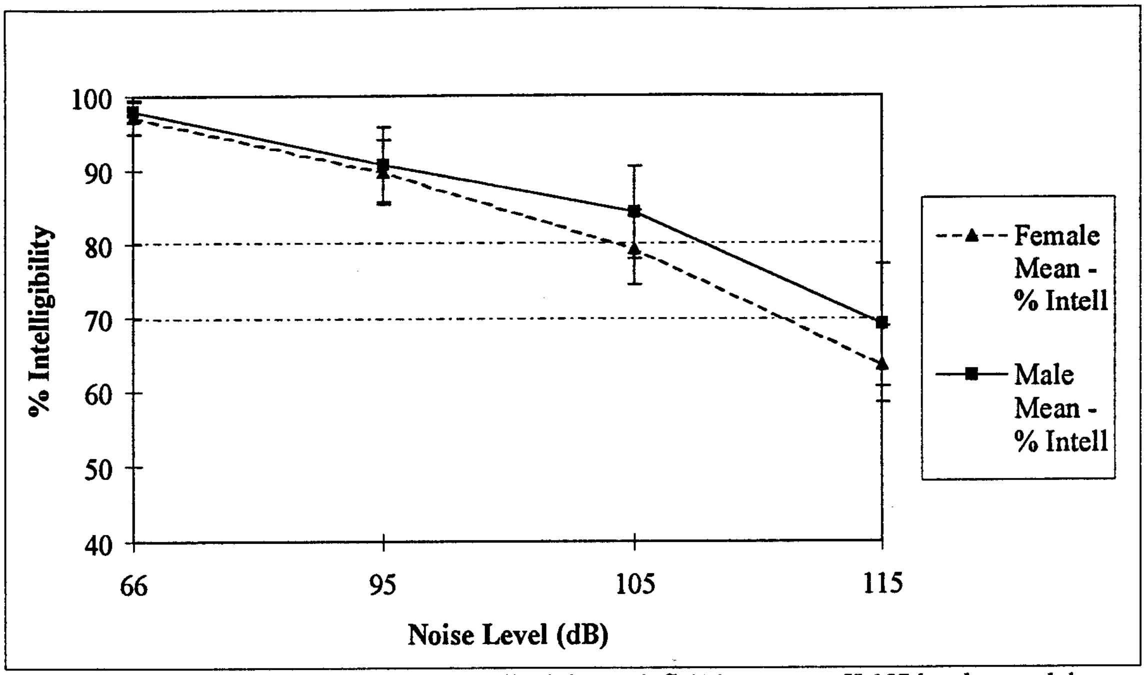


Figure 11: Phase II - Male versus female intelligibility with C-141 spectrum, H-157 headset, and the M-162 microphone.

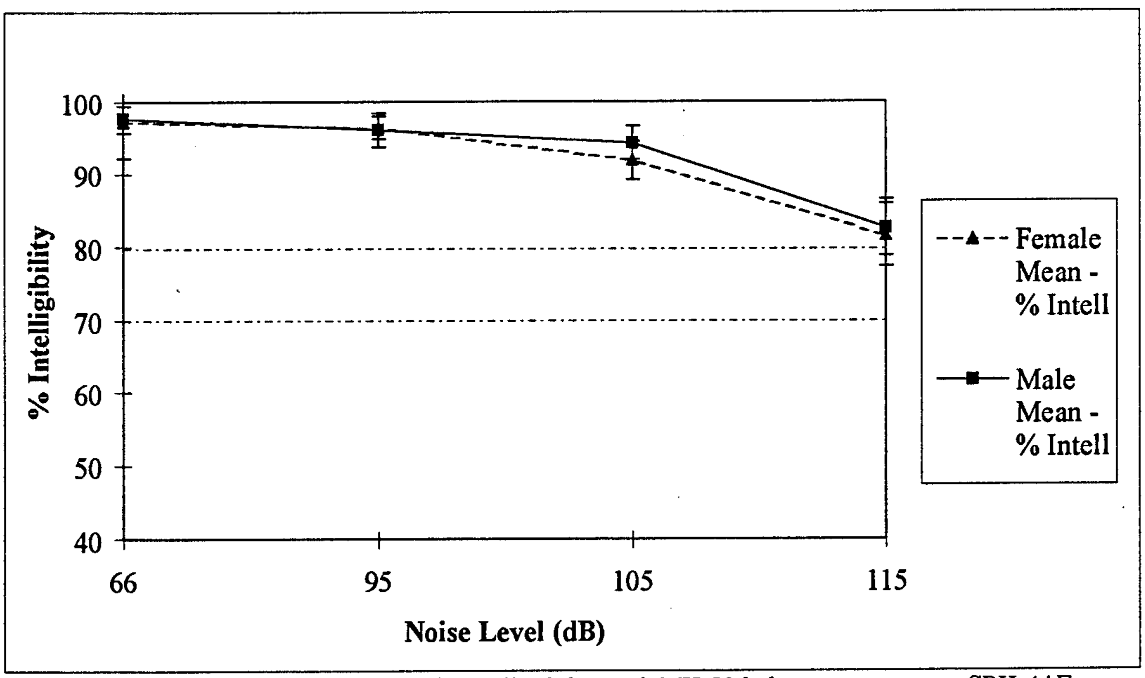


Figure 12: Phase II - Male versus female intelligibility with MH-53 helicopter spectrum, SPH-4AF helmet, and the M-162 microphone.

	66 dB	95 dB	105 dB	115 dB
Female - % avg. intelligibility ± standard deviation	96.5 ± 1.9	96.5 ± 2.2	90.6 ± 5.3	84.3 ± 3.9
Male - % avg. intelligibility ± standard deviation	97.1 ± 1.9	97.0 ± 2.6	94.4 ± 4.4	88.1 ± 4.6
Difference in Means	-0.6	-0.5	-3.8	-3.8
T-score	-0.73	-0.53	-1.73	-1.97

Table 6: Phase II - Male versus female intelligibility with C-130 spectrum, H-157 headset, and the M-162 microphone.

	66 dB	95 dB	105 dB	115 dB
Female - % avg. intelligibility ± standard deviation	97.1 ± 2.4	89.7 ± 4.4	79.4 ± 5.1	63.6 ± 5.1
Male - % avg. intelligibility ± standard deviation	97.9 ±1.4	90.8 ± 5.0	84.3 ± 6.4	68.9 ± 8.3
Difference in Means	-0.8	-1.1	-4.9	-5.3
T-score	-0.94	-0.52	-1.89	-1.7

Table 7: Phase II - Male versus female intelligibility with C-141 spectrum, H-157 headset, and the M-162 microphone.

	66 dB	95 dB	105 dB	115 dB
Female - % avg. intelligibility ± standard deviation	97.2 ± 1.7	96.5 ± 2.6	91.9 ± 4.3	81.6 ± 5.5
Male - % avg. intelligibility ± standard deviation	97.6 ± 1.9	96.1 ± 2.4	94.3 ± 2.3	82.6 ± 3.8
Difference in Means	-0.4	0.4	-2.4	-1.0
T-score	-0.54	0.35	-1.56	-0.45

Table 8: Phase II - Male versus female intelligibility with MH-53 helicopter spectrum, SPH-4AF helmet, and the M-162 microphone.

Performance of the M-87 and the M-162 microphones with female produced speech is shown in Figures 13 through 15 and Tables 9 through 11 and for male speech in Figures 16 through 18 and Tables 12 through 14. Mean speech intelligibility with the M-162 is better than with the M-87 for all aircraft and all levels of noise. Female and male speech perception with the M-162 is acceptable in all conditions except for the C-141B at 115 dB level of noise. Female speech performance with the M-87 is marginal, and with the M-162 is acceptable in the C-130E at the 115 dB level of noise. Both microphones are unacceptable for the C-141 115 dB noise condition and the M-87 is marginal in the MH-53 helicopter spectrum at 115 dB of noise.

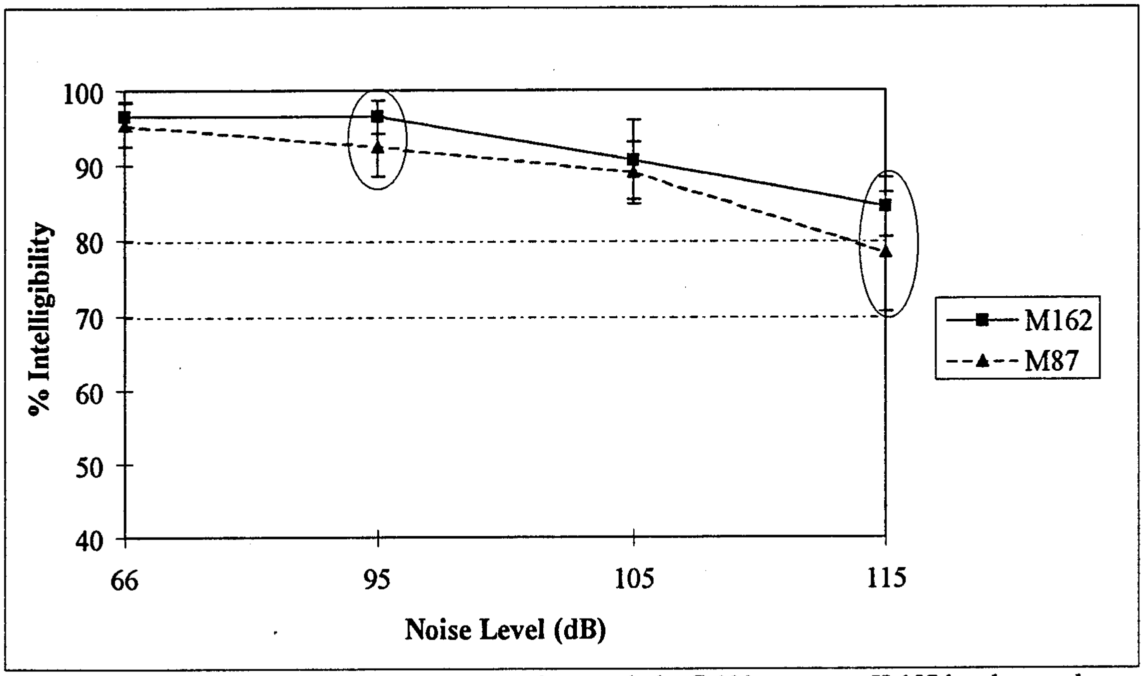


Figure 13: Phase II - M-162 versus M-87 microphone with the C-130 spectrum, H-157 headset, and female subjects.

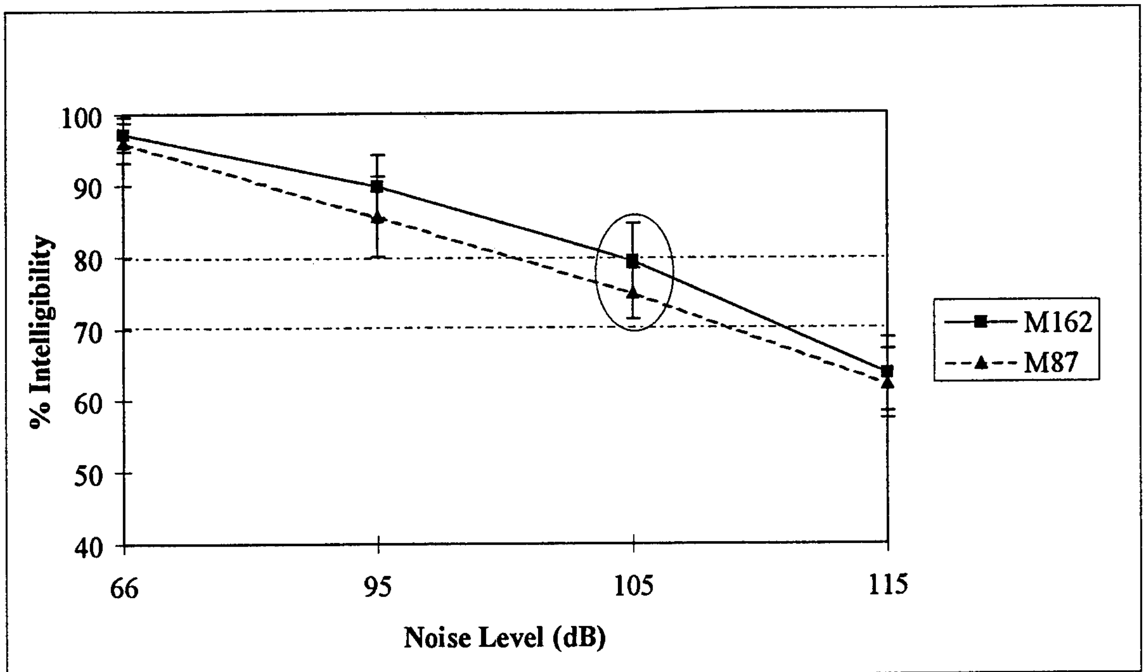


Figure 14: Phase II - M-162 versus M-87 microphone with the C-141 spectrum, H-157 headset, and female subjects.

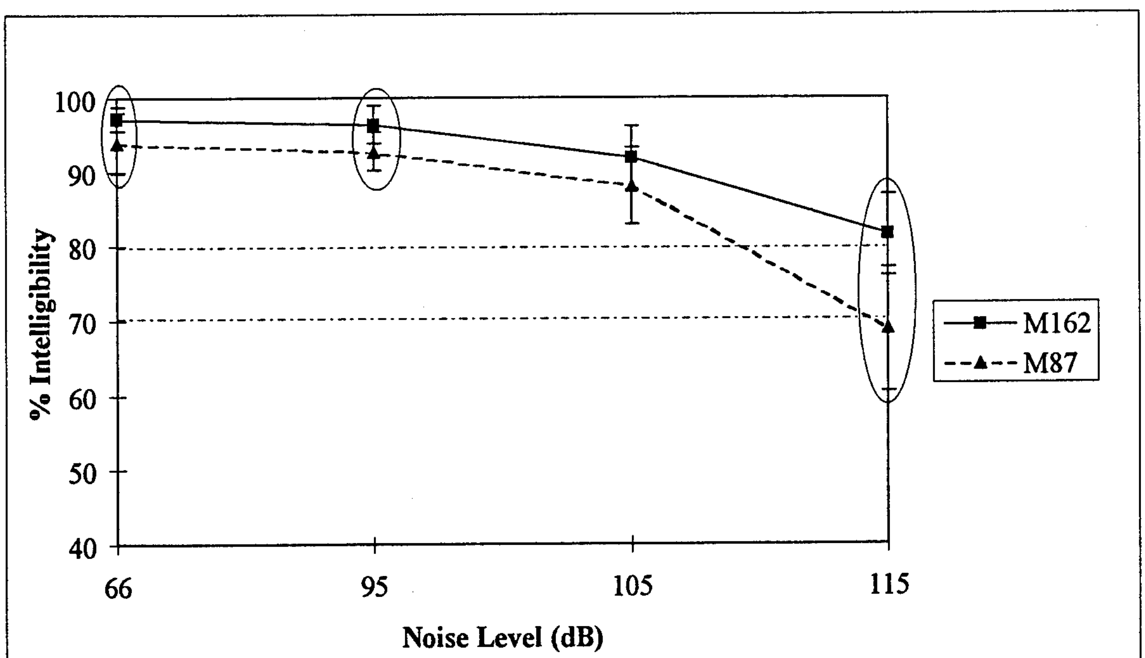


Figure 15: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and female subjects.

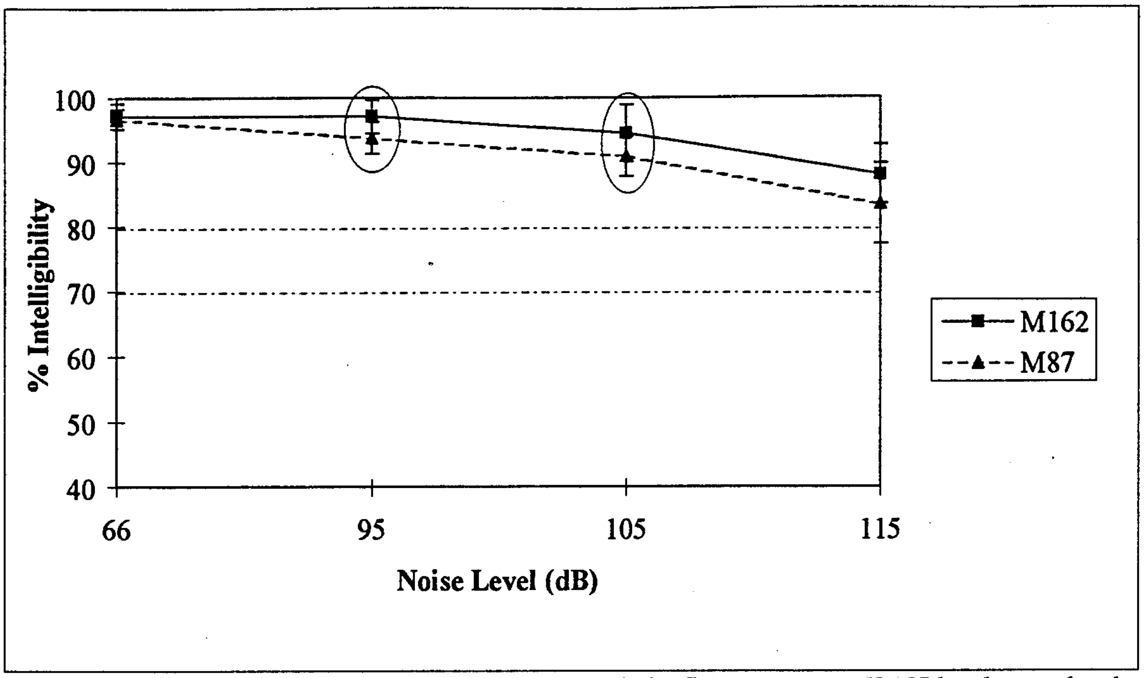


Figure 16: Phase II - M-162 versus M-87 microphone with the C-130 spectrum, H-157 headset, and male subjects.

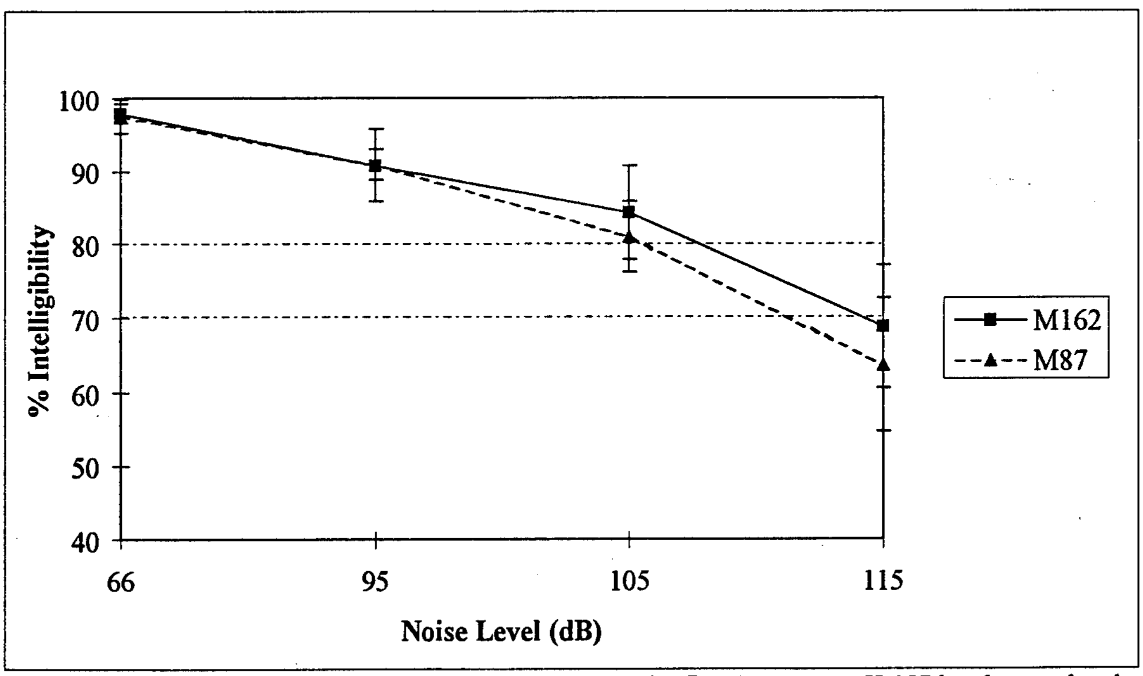


Figure 17: Phase II - M-162 versus M-87 microphone with the C-141 spectrum, H-157 headset, and male subjects.

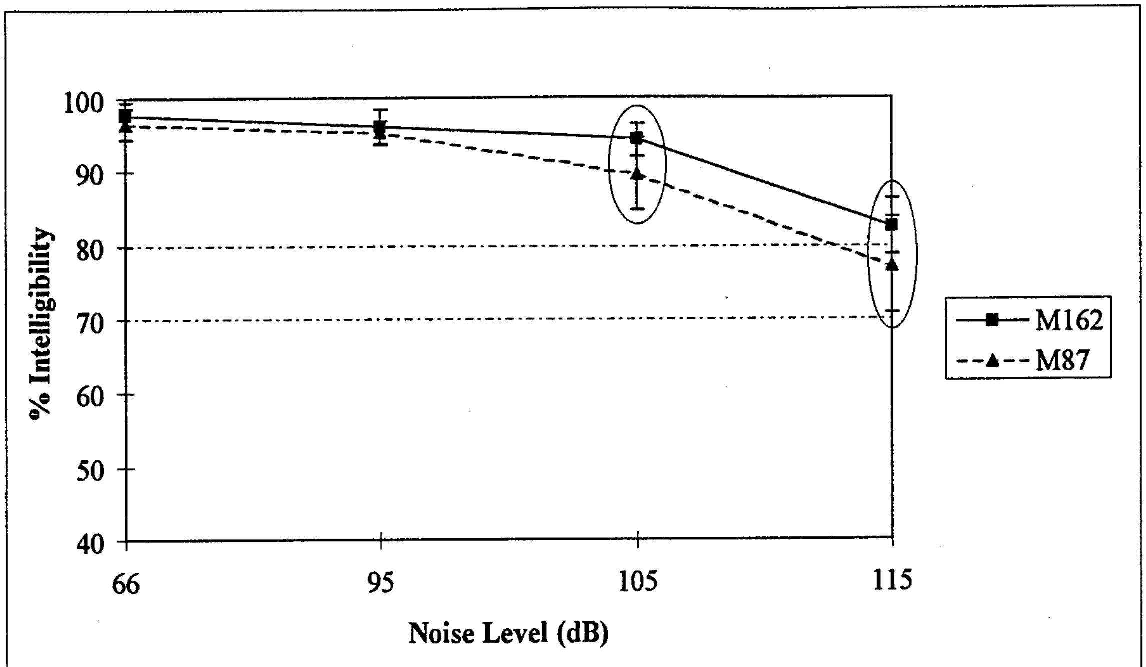


Figure 18: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and male subjects.

	66 dB	95 dB	105 dB	115 dB
M-162 - % avg. intelligibility ± standard deviation	96.5 ± 1.9	96.5 ± 2.2	90.6 ± 5.3	84.3 ± 3.9
M-87 - % avg. intelligibility ± standard deviation	95.4 ± 2.9	92.4 ± 4.0	89.0 ± 4.1	78.5 ± 7.9
Difference in Means	1.1	4.1	1.6	5.8
T-score	1.00	2.78	0.77	2.11

Table 9: Phase II - M-162 versus M-87 microphone with the C-130 spectrum, H-157 headset, and female subjects.

	66 dB	95 dB	105 dB	115 dB
M-162 - % avg. intelligibility ± standard deviation	97.1 ± 2.4	89.7 ± 4.4	79.4 ± 5.1	63.6 ± 5.1
M-87 - % avg. intelligibility ± standard deviation	96.0 ± 2.8	85.5 ± 5.6	74.9 ± 3.5	62.2 ± 4.8
Difference in Means	1.1	4.2	4.5	1.4
T-score	0.97	1.85	2.32	0.63

Table 10: Phase II - M-162 versus M-87 microphone with the C-141 spectrum, H-157 headset, and female subjects.

	66 dB	95 dB	105 dB	115 dB
M-162 - % avg. intelligibility ± standard deviation	97.2 ± 1.7	96.5 ± 2.6	91.9 ± 4.3	81.6 ± 5.5
M-87 - % avg. intelligibility ± standard deviation	93.9 ± 4.1	92.8 ± 2.6	88.2 ± 5.2	68.9 ± 8.3
Difference in Means	3.3	3.6	3.7	12.7
T-score	2.31	3.12	1.72	4.06

Table 11: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and female subjects.

	66 dB	95 dB	105 dB	115 dB
M-162 - % avg. intelligibility ± standard deviation	97.1 ± 1.9	97.0 ± 2.6	94.4 ± 4.4	88.1 ± 4.6
M-87 - % avg. intelligibility ± standard deviation	96.7 ± 1.6	93.8 ± 2.5	90.8 ± 3.1	83.6 ± 6.2
Difference in Means	0.4	3.2	3.6	4.5
T-score	0.50	2.82	2.12	1.85

Table 12: Phase II - M-162 versus M-87 microphone with the C-130 spectrum, H-157 headset, and male subjects.

	66 dB	95 dB	105 dB	115 dB
M-162 - % avg. intelligibility ± standard deviation	97.9 ± 1.4	90.8 ± 5.0	84.3 ± 6.4	68.9 ± 8.3
M-87 - % avg. intelligibility ± standard deviation	97.5 ± 2.3	91.0 ± 2.1	81.1 ± 4.9	63.7 ± 9.0
Difference in Means	0.4	-0.2	3.2	5.2
T-score	0.57	-0.10	1.27	1.35

Table 13: Phase II - M-162 versus M-87 microphone with the C-141 spectrum, H-157 headset, and male subjects.

	66 dB	95 dB	105 dB	115 dB
M-162 - % avg. intelligibility ± standard deviation	97.6 ± 1.9	96.1 ± 2.4	94.3 ± 2.3	82.6 ± 3.8
M-87 - % avg. intelligibility ± standard deviation	96.5 ± 2.1	95.3 ± 1.6	89.7 ± 4.8	77.3 ± 6.5
Difference in Means	1.1	0.8	4.6	5.3
T-score	1.24	0.79	$\boxed{2.72}$	$\boxed{2.19}$

Table 14: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and male subjects.

The conditions for which the mean differences were statistically significant at the 95 percent confidence level exhibited no pattern relative to the noise conditions or subjects. The general patterns of percent correct intelligibility showed reduced intelligibility with increased level of noise. The performance of the M-162 microphone exceeded that of the M-87 microphone in all conditions by a margin of about 5 percent, except for the MH-53 noise condition of 115 dB where it was twelve percent. The t-scores for these conditions are relatively low and close to the critical t-value, and the statistical significance is influenced by the variance of the data. The t-scores decrease as the variance increases for the same N.

The Phase II data indicate that the mean female speech perception is lower than the mean male speech perception for both microphones in all conditions; however, the amount of difference is relatively small and not statistically significant. As seen in Figure 19, the intelligibility of female speech is about 12 percent better with the M-162 microphone than with the M-87. This improvement in speech intelligibility measured with the M-162 is evident in all the experimental conditions for the C-130E, the C-141B, and the MH-53 helicopter (Figures 19-21). These data suggest that the perception of both female and male speech may be improved in the three aircraft spectra at all noise conditions by replacing the M-87 microphone with the M-162 microphone.

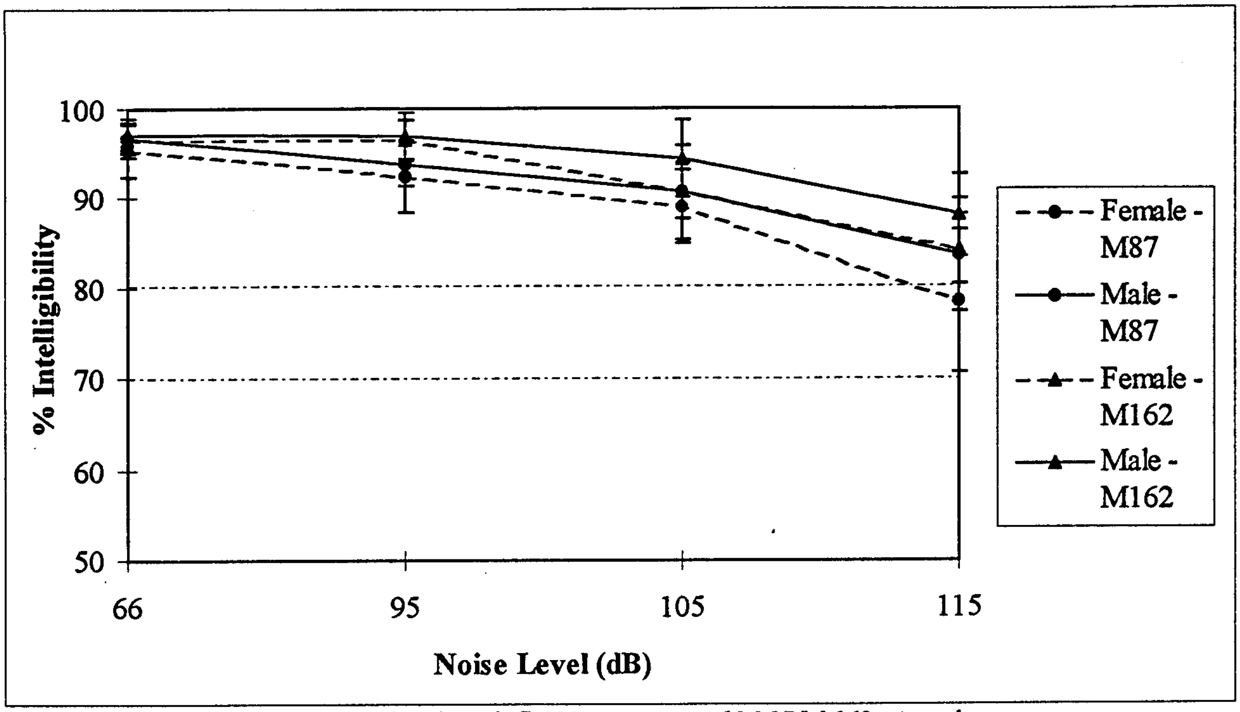


Figure 19: Phase II - Male versus female with C-130 spectrum and M-87/M-162 microphones

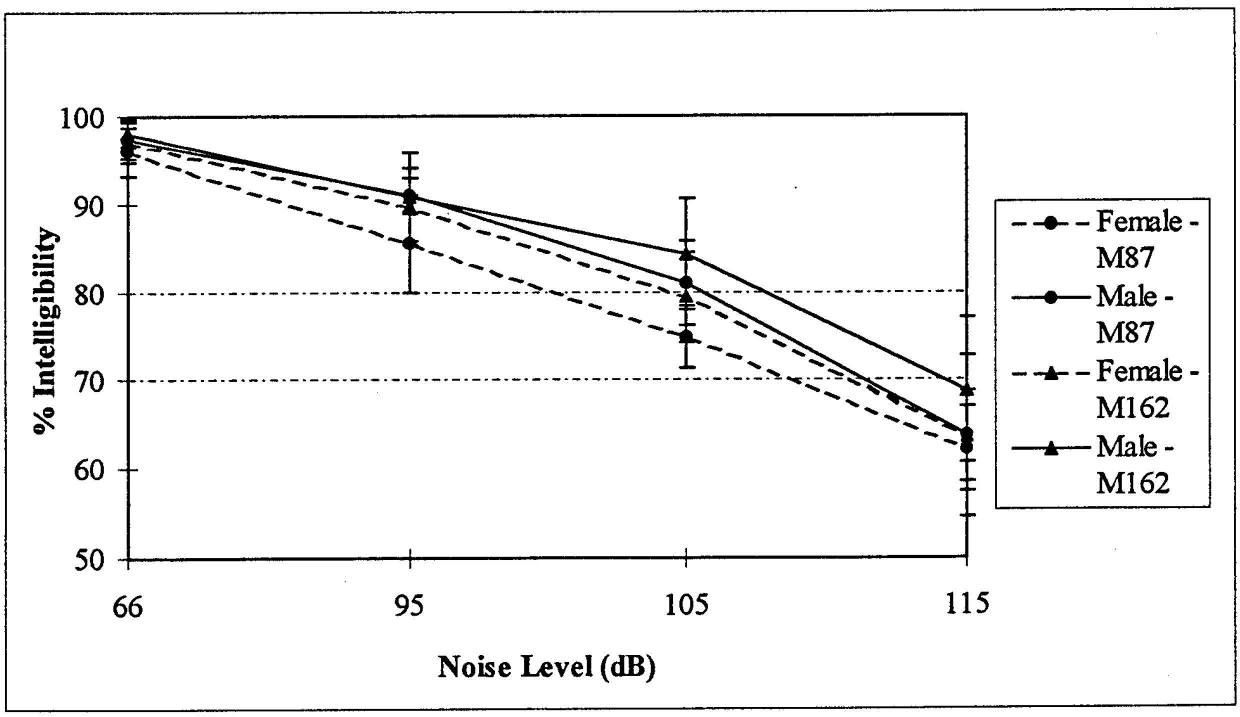


Figure 20: Phase II - Male versus female with the C-141 spectrum and M-87/M-162 microphones.

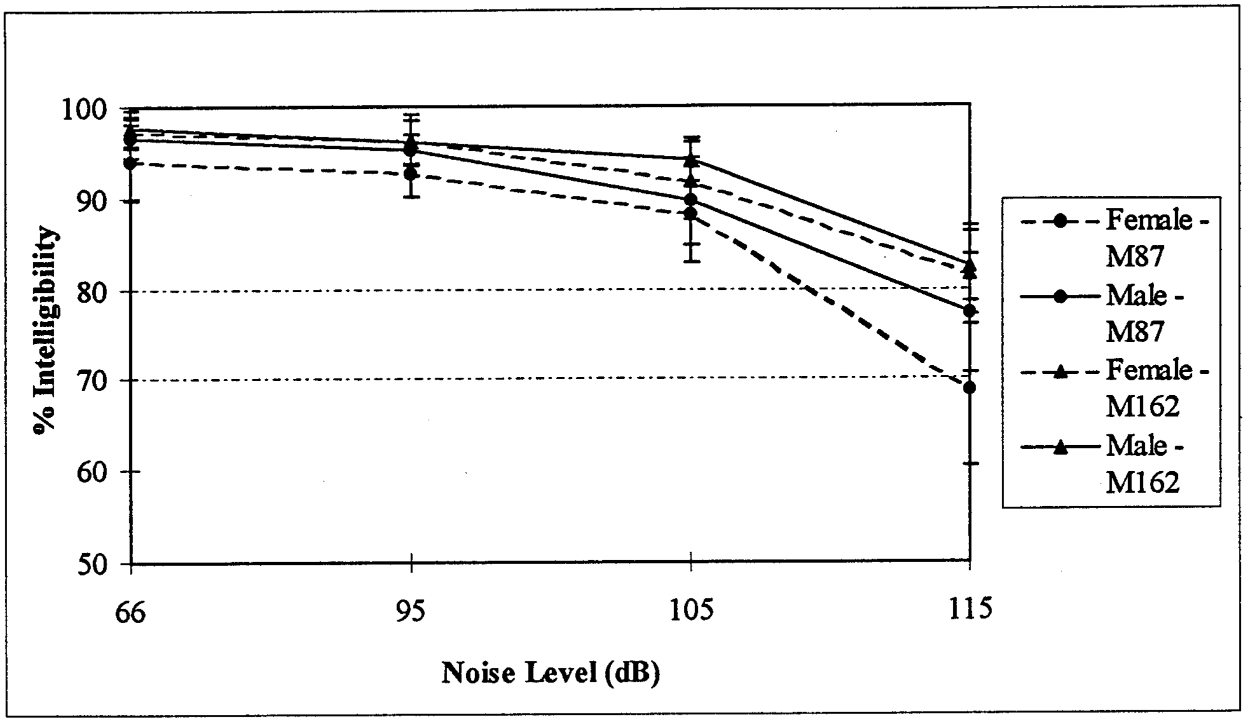


Figure 21: Phase II - Male versus female with the MH-53 spectrum and the M-87/M-162 microphones.

SUMMARY

Overall, results from both Phases I and II reveal that the mean percent correct intelligibility of female produced speech was lower than the mean intelligibility of male produced speech by as much as ten percent, and more. The general trend indicated that the amount of the difference between the male and female speech increased as the level of the noise condition increased. The maximum effect usually, but not always, occurred at the condition of highest level of noise. The data also indicated a number of conditions for which the average differences between the female and male speech were statistically significant at the 95 percent level of confidence. These conditions of statistical significance did not follow the trend displayed by the decreasing speech communication effectiveness with increasing level of noise, but were somewhat random in occurrence. However, each one of those conditions does verify that the female speech is less intelligible than the male speech at least 95 percent of the time and that the data demonstrating the poorer female speech perception is real. The mean differences for the statistically significant conditions ranged from about three to six percent; however, other conditions with mean differences within this range, and even higher, were not statistically significant. The differences between these averages for both sets of data are relatively small and represent two and three word errors in an MRT list of 50 words. Observation of the percent correct intelligibility data reveals very few situations where a three to six percent difference is meaningful in an operational situation.

Perception of speech in the operational situation was also evaluated using the performance criteria or biocommunications guidelines described earlier. Percent correct intelligibility is compared to benchmark values in the regions below, between, and above the 70 and 80 percent correct intelligibility levels. Laboratory performance that exceeds 80 percent correct translates to acceptable operational performance and that below 70 percent to unacceptability. Performance in the marginal area between the 70 and 80 percent values means that operational performance may or may not be acceptable, depending on the specific conditions and requirements. The laboratory values that are close to the 70 percent and 80 percent (which are not pass-fail values) are in the fringe areas and may require more information than just the intelligibility scores for a confident estimation of the real world performance. The overall speech intelligibility performance for the conditions in Phases I and II are summarized relative to the performance criteria in Table 15. The data are coded such that estimated operational acceptability is equal to + (80 percent and above), marginal acceptability is equal to ±, and operational unacceptability is - (69 percent and below).

			Level of Aircraft Noise (dB)			
Phase	Aircraft	Gender	75	95	105	115
I	C-130	M	+	+	+	+
		F	+	+	±	±
	C-141	M	+	+	+	-
		F	+	+	±	-
	F-15	M	+	+	+	±
		F	+	+	+	-
	MH-53	M	+	+	+	±
		F	+	+	+	-
II	C-130	M	+	+	+	+
		F	+	+	+	+
	C-141	M	+	+	+	-
		F	+	+	±	-
	MH-53	M	+	+	+	+
		F	+	+	+	+

Table 15: Summary table of average percent correct intelligibility of female and male speech in Phases I and II, evaluated by performance criteria. Acceptable = +, marginal = ±, and unacceptable = –.

INTERIM CONCLUSIONS

1. Mean female speech is less intelligible than mean male speech in all experimental conditions measured in Phases I and II. However, the differences in intelligibility are not always statistically significant and may not be meaningful in operational situations.

2. These mean differences in intelligibility between male and female speech tend to increase as the levels of the noises increase (from 66 dB to 115 dB).

3. The statistically significant differences between mean intelligibility of male and female speech occurred in a somewhat random fashion. No patterns emerged that were associated with the experimental variables.

4. Examination of the four aircraft cockpit noise spectra at cruise (fixed wing) and hover (rotary wing) indicates that female speech is five to seven percent less intelligible than male speech during cruise. However, both types of speech are acceptable in the C-130E, C-141B, and MH-53. Male speech would be marginal and female speech unacceptable in the F-15A noise at the 115 dB level.

5. Female speech is unacceptable in the 115 dB noise of the C-141B, F-15A, and MH-53 and is marginal in the C-130E, 115 dB noise and the F-15A, 105 dB level of noise. Male speech is unacceptable in the 115 dB noise of the C-141B and marginal in the 115 dB noise of the F-15A.

6. Using the M-162 noise-cancelling microphone, both male and female (less than male) speech intelligibility were acceptable in all C-130E and MH-53 noise environments and both were unacceptable in the C-141B spectrum at 115 dB.

7. Speech intelligibility of both female and male speech with the M-162 microphone was as much as 12 percent better than with the M-87 microphone. The greatest improvements occurred in the highest levels of noise.

INTERIM RECOMMENDATIONS

Initial interpretations of the data suggest that the following actions might alleviate the voice communications deficiencies identified in the first two phases of this study. These recommendations can be validated with additional experimentation in the unique voice communications emulation facilities of the Bioacoustics and Biocommunications Branch.

1. Replace the M-87 noise-cancelling microphones with the M-162 noise-cancelling microphones. This would immediately bring the perception of female speech to the current perception level of male speech using the M-87 microphone. The speech intelligibility of the male speech would also experience comparable improvements.

2. Provide headsets and helmets with appropriate active noise reduction (ANR) capability. Because of our extensive experience with ANR technology, we predict that this technology would improve the speech intelligibility in the cockpit environment. The Air Force has developed a flight-worthy circumaural headset technology that will undergo Operational Test and Evaluation in the near term.

3. Complete development of a lightweight ANR headset for non-flight-helmet applications such as C-130E and C-140 type aircraft. Because of our experience, we predict that this new technology would improve communications in these aircraft also.

REFERENCES

1. ANSI S12.6-1984 (R 1990), American National Standard Method for the Measurement of Real-Ear Attenuation of Hearing Protectors.

2. ANSI S3.2-1989 (ASA 85), American National Standard Method for Measuring the Intelligibility of Speech over Communications Systems

3. Backs, R. W., and Walrath, L. C., "(A) Heart Rate and Auditory Workload During Noise Stress," Proceedings of 6th International Symposium on Aviation Psychology, Vol 2 (A92-44901-19-53), Ohio State University, Columbus OH, 1991.

4. Box, George E. P., Hunter, William G., and Hunter, J. Stuart, Statistics for Experimenters, John Wiley & Sons, Inc., 1978.

5. C-130E In-Flight Crew Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 38, WPAFB OH, September 1975.

6. C-141 In-Flight Crew/Passenger Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 126, WPAFB OH, June 1980

7. F-15A In-Flight Crew Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 127, WPAFB OH, August 1979.

8. Fletcher, Harvey, Speech and Hearing in Communication, D. Van Nostrand Company, Inc., 1953.

9. Freedman, Jay and Rumbaugh, William A. "Accuracy and Speech of Response to Different Voice Types in A Cockpit Voice Warning System," Air Force Institute of Technology, LSSR 89-83, Wright-Patterson AFB OH, 1983.

10. HH-53C In-Flight Crew Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 51, WPAFB OH, October 1975.

11. Holden, James M. And Vensko, George, Voice R&D Progress in the Helicopter Environment, Proceedings of Speech Tech 89, Media Dimensions, New York NY, 1989.

12. House, A. S., Williams, Carl E., Hecker, Michael H. L., and Kryter, Karl D., "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set," Journal of the Acoustical Society of America, 37 (1965), 158-166.

13. McKinley, Richard L., "Voice Communication Research and Evaluation System," Aerospace Medical Research Laboratory, AMRL-TR-80-25, 1980

14. McKinley, Richard L., Nixon, Charles W., and Moore, Thomas J., "Voice Communication Capability of Selected In-Flight Headgear Devices," AGARD Proceedings, Soesterberg, Netherlands, March-April 1981.

15. Moore, Thomas J., Nixon, Charles W., and McKinley Richard L., "Comparative Intelligibility of Speech Materials Processed by Standard Air Force Communications Systems in the Presence of Simulated Cockpit Noise,

16. Nixon, Charles W., Moore, Thomas, J., and McKinley, Richard L., "Increase in Jammed Word Intelligibility Due to Training of Listeners," Journal of Aviation and Space Medicine, AAFMRL-81-64, 1982.

17. Nixon, Charles W. and McKinley, Richard L., "Intelligibility in Noise of Three LPC Voice Channels with ANR Headsets," AAMRL-TR-88-063, Wright-Patterson AFB OH, 45433, November 1988.

18. Nixon, C. W. And McKinley , R. L., "LPC-10 Intelligibility of Oxygen Masks and Microphones in Noise," Armstrong Aerospace Medical Research Laboratory, AAMRL-TR-88-048, November 1988.

19. Nixon, Charles W., "Voice Communications and Positive Pressure Breathing," AFAMRL-TR-84-009, Wright-Patterson AFB OH, 45433, January 1984.

20. Prinzo, O. Veronika and Britton, Thomas W., "ATC/Pilot Voice Communications - A Survey of the Literature," Civil Aeromedical Institute, FAA, Oklahoma City OK, 73125, (p. 6), November 1993.

21. Simpson, Carol, "Evaluation of Speech Recognizers for use in Advanced Combat Helicopter Crew Station Research and Development," Ames Research Center, NASA Contractor Report 177547, US Army Aviation Systems Command, TM-90-A-001, Moffet Field, CA 94035, March 1990.

22. Simpson, Carol A., Marchionda-Frost, Kristine, and Navarro, Teresa, "Comparison of Voice Types for Helicopter Voice Warning Systems, NASA Contract NAS2-11341, Report 841611, Ames Research Center, Moffett Field CA.

23. Williamson, David T., and Curry, David G., Speech Recognition Performance Evaluation in Simulated Cockpit Noise," Proceedings of Speech Tech 84, Media Dimensions, New York NY, 1984.

24. Werkowitz, Eric, "Speech Recognition in the Tactical Environment: The AFTI-F-16 Voice Command, Proceedings of Speech Tech 84, Media Dimensions, New York NY, 1984.

APPENDIX A

		Frequency (Hz)						
Microphone Type	Mic Number	125	250	500	1000	2000	4000	8000
Brüel & Kjær	4134	1.60	0.40	0.67	0.45	1.30	0.81	1.06
M-87	1	2.12	4.00	7.80	8.97	9.63	5.23	1.23
M-87	2	1.43	2.71	6.36	9.31	7.29	4.08	1.22
M-87	3	1.63	3.60	8.46	9.96	7.60	4.78	1.09
M-87	4	1.46	2.99	9.00	6.76	4.93	5.30	0.78
M-87	5	2.14	4.32	9.09	8.68	9.72	3.20	1.58
M-87	6	1.63	3.11	6.69	11.52	13.70	4.10	1.74
M-87	7	1.20	2.14	4.63	9.68	5.89	5.38	2.00
M-87	8	2.47	4.96	9.46	10.75	11.93	5.11	1.29
M-87	9	2.47	4.83	9.55	8.97	9.44	5.14	1.64
M-87	10	1.52	1.22	6.33	9.98	9.97	4.74	2.53

Table 16: M-87 microphone calibration data in volts rms.

		Frequency (Hz)						
Microphone Type	Mic Number	125	250	500	1000	2000	4000	8000
Brüel & Kjær	4134	1.60	0.39	0.66	0.44	1.30	0.80	1.05
M-162	11	0.83	0.86	0.85	0.81	0.87	0.56	0.31
M-162	12	0.69	0.71	0.71	0.65	0.72	0.49	0.44
M-162	13	0.96	0.87	0.90	0.75	0.90	0.56	0.31
M-162	14	0.78	0.69	0.71	0.61	0.70	0.40	0.37
M-162	15	0.68	0.73	0.77	0.68	0.85	0.58	0.39
M-162	16	0.70	0.65	0.70	0.63	0.77	0.80	0.27
M-162	17	0.69	0.62	0.66	0.57	0.69	0.41	0.28
M-162	18	0.63	0.55	0.89	0.49	0.58	0.34	0.40
M-162	19	0.96	0.89	0.94	0.81	1.00	0.62	0.30
M-162	20	0.86	0.81	0.88	0.78	0.98	0.64	0.29

Table 17: M-162 microphone calibration data in volts rms.

		Frequency (Hz)						
Microphone Type	Mic Number	125	250	500	1000	2000	4000	8000
Brüel & Kjær	4134	1.60	0.39	0.66	0.44	1.30	0.80	1.05
M-169	21	2.28	4.26	8.07	11.67	12.11	10.01	3.87
M-169	22	2.15	3.99	7.25	10.75	11.73	10.41	5.76
M-169	23	2.35	4.35	8.00	11.76	12.00	10.28	5.56
M-169	24	1.96	3.61	6.85	10.79	11.86	10.23	6.26
M-169	25	1.97	3.60	6.50	9.40	10.47	9.76	5.73
M-169	26	1.88	3.37	6.25	9.32	10.32	10.40	5.88
M-169	27	2.12	3.94	7.46	11.34	11.91	9.87	5.27
M-169	28	2.09	3.91	7.24	9.79	10.50	8.71	6.10
M-169	29	2.10	3.66	6.03	8.13	9.54	9.45	4.48
M-169	30	2.03	3.70	6.58	9.93	11.42	10.63	6.66

Table 18: M-169 microphone calibration data in volts rms.

APPENDIX B

Headset/ Helmet Type	Average (m) or one standard deviation (1σ)	Frequency (Hz) 125	250	500	1000	2000	3150	4000	6300	8000
HGU-55P	m	8	2	10	23	37	42	45	47	47
	1σ	4.3	4.0	4.6	5.3	5.8	4.8	5.2	6.9	7.1
H-157A	m	10	12	18	32	38	39	37	37	35
	1σ	2.6	2.9	3.6	6.2	4.3	4.9	6.1	7.3	6.0
SPH-4AF	m	14	13	24	37	38	40	40	45	43
	1σ	2.7	2.2	2.2	5.3	2.6	4.1	4.3	5.0	4.8

Table 19: Average and standard deviation of headset/helmet attenuation (sound pressure level, dB).

TABLE OF FIGURES

TABLE OF TABLES